Nuclear Safety
NEA/CSNI/R(2015)1
March 2015
www.oecd-nea.org

Establishing the Appropriate Attributes in Current Human Reliability Assessment Techniques for Nuclear Safety

OECD
BETTER POLICIES FOR BETTER LIVES

NEA
NUCLEAR ENERGY AGENCY

Organisation de Coopération et de Développement Économiques
Organisation for Economic Co-operation and Development

English text only

NEA/CSNI/R(2015)1
Unclassified

Establishing the Appropriate Attributes in Current Human Reliability Assessment Techniques for Nuclear Safety

English text only

COMMITTEE ON THE SAFETY OF NUCLEAR INSTALLATIONS

The NEA Committee on the Safety of Nuclear Installations (CSNI) is an international committee made of senior scientists and engineers, with broad responsibilities for safety technology and research programmes, as well as representatives from regulatory authorities. It was set up in 1973 to develop and co-ordinate the activities of the NEA concerning the technical aspects of the design, construction and operation of nuclear installations insofar as they affect the safety of such installations.

The committee's purpose is to foster international co-operation in nuclear safety amongst the NEA member countries. The CSNI's main tasks are to exchange technical information and to promote collaboration between research, development, engineering and regulatory organisations; to review operating experience and the state of knowledge on selected topics of nuclear safety technology and safety assessment; to initiate and conduct programmes to overcome discrepancies, develop improvements and research consensus on technical issues; and to promote the co-ordination of work that serves to maintain competence in nuclear safety matters, including the establishment of joint undertakings.

The clear priority of the committee is on the safety of nuclear installations and the design and construction of new reactors and installations. For advanced reactor designs the committee provides a forum for improving safety related knowledge and a vehicle for joint research.

In implementing its programme, the CSNI establishes co-operative mechanisms with the NEA's Committee on Nuclear Regulatory Activities (CNRA) which is responsible for the programme of the Agency concerning the regulation, licensing and inspection of nuclear installations with regard to safety. It also co-operates with the other NEA's Standing Committees as well as with key international organizations (e.g., the IAEA) on matters of common interest.

FOREWORD

The work described in this report was conducted as a joint task under the CSNI Working Groups on Human and Organisational Factors (WGHOF) and Risk Assessment (WGRISK). The task has two primary purposes: to identify a set of desirable attributes for current HRA techniques used in nuclear risk assessment and to evaluate a set of HRA techniques used in OECD member countries against these attributes. The aim is to provide information that will support regulators and operators of nuclear facilities when making judgements about the appropriateness of HRA methods for conducting assessments in support of Probabilistic Safety Assessments (PSA).

Both WGHOF and WGRISK have provided active forums for information exchange on the topic of Human Reliability Analysis (HRA) and have engaged in previous joint projects on this topic e.g. on simulator studies for HRA purposes [See NEA/CSNI/R(2012)1]. WGRISK's past efforts have addressed HRA practice and data issues [see NEA/CSNI/R(98)1] and the development of methods for Errors of Commission [see NEA/CSNI/R(2000)17 and NEA/CSNI/R(2002)3]. In 2004, the working group issued a Topical Opinion Paper on HRA [CSNI Technical Opinion Paper No. 4] that identified the scarcity of empirical human performance data as a significant challenge. In 2008, WGRISK issued a report [NEA/CSNI/R(2008)9] addressing the feasibility of a joint international effort on HRA data.

This work represents the collective effort of the task group all of whom provided valuable time and considerable knowledge toward its production.

Task steering group

Jane Bowie	ONR	United Kingdom
Gary Munley	ONR	United Kingdom
Vinh Dang	PSI	Switzerland
John Wreathall	(Consultant)	United States

Task group members and OECD staff supporting the task

* Attribute generation phase only
\+ Method evaluation phase only

Andreas Bye	IFE OECD Halden Reactor Project	Norway
Susan Cooper+	USNRC	United States
Veronique Fauchille	IRSN	France
Jean Yves Fiset+	CNRC	Canada
Monica Haage	IAEA	Austria
Gunnar Johanson	ES Konsult	Sweden
Won Dae Jung+	KAERI	Republic of Korea
Jaewhan Kim	KAERI	Republic of Korea
Jan Kubicek	Nuclear Institute Rez	Czech Republic
Pierre Le Bot*	EDF	France
Seung Jung Lee+	KAERI	Republic of Korea
Julie Marble*	USNRC	United States
Helene Pesme*	EDF	France
Sean Peters*	USNRC	United States
Wolfgang Preischl	GRS	Germany
Alice Salway*	CNSC	Canada
Jing Xing+	USNRC	United States

OECD Nuclear Energy Agency

Abdallah Amri
Greg Lamarre
Andrew White

TABLE OF CONTENTS

Foreword 3

Executive summary 7

1. Introduction 11
1.1 Purpose 11
1.2 Background 11
1.3 State of the art in human reliability analysis 12
1.4 Scope 12

2. Study phases and steps 13
2.1 Overview 13
2.2 Phase 1: Development of attributes and evaluation method 13
2.2.1 Generation of attributes 13
2.2.2 Applying the attributes 14
2.2.3 Attribute grouping 15
2.2.4 Attribute importance weighting 15
2.2.5 Selection of HRA methods to be evaluated 17
2.3 Phase 2: Conduct of HRA method evaluations 17
2.3.1 Input from method developers 17
2.3.2 Method evaluation process 18
2.3.3 Evaluation team assessment 18
2.3.4 Task group assessment 18
2.3.5 Consistency review 19
2.3.6 Note on the evaluation of the HuRECA method 19

3. The attributes 21
3.1 Construct validity 21
3.2 Content validity 23
3.3 Empirical validity 27
3.4 Reliability 28
3.5 Usability 30

4. Results 33
4.1 Introduction 33
4.2 Technique for human error rate prediction (THERP) 33
4.3 ASEP HRA method 34
4.4 Enhanced Bayesian THERP method 35
4.5 A technique for human event analysis (ATHEANA) 36
4.6 MERMOS 37
4.7 NARA 38
4.8 Standardised plant analysis risk human reliability analysis methodology (SPAR-H) 39
4.9 HCR/ORE & CBDT methods 39
4.10 Cognitive reliability and error analysis method (CREAM) 40
4.11 Failure likelihood index method (FLIM) 41

4.12 HuRECA 42
4.13 Summary of results 43

5. Discussion 45
5.1 Introduction 45
5.2 Construct validity 45
5.3 Content validity 47
5.4 Empirical validity 50
5.5 Reliability 51
5.6 Usability 52
5.7 Limitations of the study 53
5.8 Areas for further research 54

6. Conclusion 55

7. References 57

Appendices
1. The method evaluation scale 61
2. Attribute evaluations for each method 71
A2.1 Attribute evaluations – THERP & ASEP 71
A2.2 Attribute evaluations – Enhanced Bayesian THERP 87
A2.3 Attribute evaluations – ATHEANA 103
A2.4 Attribute evaluations –MERMOS 119
A2.5 Attribute evaluations – NARA 136
A2.6 Attribute evaluations – SPAR-H 153
A2.7 Attribute evaluations –HCR/ORE & CBDT 170
A2.8 Attribute evaluations – CREAM 187
A2.9 Attribute evaluations – FLIM 204
A2.10 Attribute evaluations – HuRECA 222
3. Method developers 239
A3.1 Developer's comments on method evaluation – Enhanced Bayesian THERP 240
A3.2 Developer's comments on method evaluation – NARA 256
A3.3 Developer's comments on method evaluation – HCR/ORE & CBDT 273

EXECUTIVE SUMMARY

This report presents the results of a joint OECD NEA CSNI WGRISK/WGHOF task to identify and define desirable attributes of Human Reliability Assessment (HRA) methods, and to evaluate a range of HRA methods used in OECD member countries against those attributes.

The study did not set out to recommend or promote the use of any particular HRA method nor does it aim to score or rank the methods. Rather the study aims to identify the strengths and limitations of new and commonly used methods, to aid those responsible for production of HRAs in selecting appropriate tools for specific HRA applications, and to assist regulators when making judgements on the appropriateness of the application of an HRA technique within nuclear-related probabilistic safety assessments. Thus, the evaluations of the methods on the individual attributes are intended to inform a decision-maker, who will need to identify the most important attributes for the application under consideration.

The project was undertaken by a task group comprised of international HRA experts. The first phase of the project was to develop a set of attributes considered important for any HRA method aimed at providing Human Error Probabilities (HEPs) for use in probabilistic safety assessments (PSAs). The second phase involved small teams of experts using the attributes to evaluate a set of HRA methods against them. A total of twenty attributes were developed and grouped into five categories:

- Construct validity – a measure of the internal validity of the method that assesses the extent to which the HRA method measures or assesses what it claims to, and is consistent with an underlying theoretical model or dataset.
- Content validity – a second measure of internal validity that assesses if the HRA method measures or assesses important determinants of human reliability.
- Empirical validity – a measure of the extent to which numerical outputs from an HRA method have been demonstrated to correlate with other sources of human reliability data.
- Reliability – a measure of extent to which an HRA method produces consistent quantitative and qualitative output and the extent to which the derivation of the output can be traced and therefore verified by a reviewer.
- Usability – a measure of the extent to which an HRA method provides clear guidance for its application, useable outputs and the amount of resource required for its implementation.

For each attribute, the experts developed consensus ratings to reflect its importance from the perspectives of two groups: potential users of HRA methods (such as utility staff) and reviewers of their results (such as regulators). These ratings help to distinguish essential attributes from desirable but less critical ones.

To select methods for evaluation in the second phase, an initial poll of task group members was undertaken to identify the HRA methods currently in use in member countries, and those for which there is interest and potential for application. From this poll, twelve methods were selected. In two cases, the THERP and ASEP, and HCR/ORE and CBDT these methods were reviewed under a single evaluation as the methods are often used together in single assessment. The methods evaluated were:

- Technique for Human Error Rate Prediction (THERP) Family – comprising THERP + Accident Sequence Evaluation Programme (ASEP).

- Enhanced Bayesian THERP (EBT).
- A Technique for Human Event Analysis (ATHEANA).
- Méthode d'Evaluation de la Réalisation des Missions Opérateurs pour la Sûreté (MERMOS).
- Nuclear Action Reliability Assessment (NARA).
- Simplified Plant Analysis Risk – Human Reliability Analysis (SPAR-H).
- Human Cognitive Reliability (HCR)/Operator Reliability Experiments (ORE) and Cause Based Decision Tree (CBDT) Methods.
- Cognitive Reliability and Error Analysis Method (CREAM).
- Failure Likelihood Index Methodology (FLIM).
- Human Reliability Evaluator for Control Room Actions (HuRECA).

Evaluation of the individual HRA methods selected for assessment in relation to the identified desirable attributes of HRA techniques was conducted in a number of stages. Before undertaking the evaluation of the method against the attributes, a principal developer of each method was contacted and invited to provide information on how the particular method addressed each attribute and where in the method's documentation such evidence could be found. The aim was to allow method developers to identify the best evidence available in relation to each of the methods in order that the task group could provide the most accurate evaluation of the method and to increase the efficiency of the method evaluations. Responses were received from developers of five methods (ATHEANA, Enhanced Bayesian THERP, HCR/ORE & CBDT, NARA and SPAR-H).

The method evaluation stages undertaken by the task group comprised:

- Evaluation Team assessment: For most method reviews, separate evaluation teams comprised of a minimum of a lead and second reviewer performed the complete evaluation of each method against the twenty attributes. Once an evaluation team was content with their evaluation of a method these were made available for the remaining members of the task group for review via an OECD web portal. In the cases of MERMOS and HuRECA the evaluations were performed slightly differently though in keeping with the above process.[1] In the evaluation of the HCR/ORE and CBDT methods, additional information was provided by the methods' developers after the initial review and therefore an updating of the review was made to take account of this information.
- Task Group assessment: In order to derive a task group consensus on the evaluation of each method, two workshops were undertaken in order to review the individual method evaluations. Once the task group had discussed a method evaluation, an updated evaluation was produced to reflect the decisions taken in the workshop. These revised evaluations were then made available via the OECD web portal for final review. (It should be noted that where consensus could not be achieved, the evaluation of the method as provided by the lead reviewer was recorded and counterarguments raised in the discussions are reported in the discussion section of this report.)
- Consistency review: Once all individual method evaluations had undergone the task group assessment process a summary table reporting the complete set of method evaluations was produced. This allowed for the comparison of individual method evaluations across each attribute. Using this summary table as an input, an exercise was undertaken to identify any apparent inconsistencies in the treatment of methods when evaluated against each attribute. This consistency review was completed as part of a workshop and was used as an opportunity to challenge any individual evaluation on the basis of a comparison against the ratings applied

1. Much of the documentation of MERMOS and all that for HuRECA was not in English so there were limitations in their reviews. For MERMOS, one reviewer reviewed the detailed French documentation while others reviewed the general information in English. HuRECA was reviewed collectively by the task group on the basis of a presentation in English made by the principal method developer, as the documentation was only available in Korean.

across the set of methods. When a change to the method evaluation was proposed, the method evaluation scale, including the justification for the evaluation, was updated to reflect the decision taken at the meeting. Each updated method evaluation was then returned to the Lead Reviewer to confirm the evaluation.

The main results of this project are comprised of the ratings of each method against each attribute, together with the written justification of the bases for the ratings. A three-point rating scale has been used. A "high" rating indicates that the requirements of an attribute have been fully or largely met for the method's intended scope of application. The "intermediate" rating indicates that a method meets some, but not all of the requirements of the attribute. The "low" rating indicates that the requirements of the attribute are not met or that no evidence is available. For a few attributes with a binary character, the intermediate rating is not used. A summary table is provided but the justifications associated with the ratings for each method are considered by the task group to be the primary product.

As well as the detailed evaluations of each of the HRA methods presented in the appendices of this report, the study generated a number of general findings which reflect on the current state of the art of HRA and areas for further development.

In the member countries as a whole, the relevant attributes for the selection and acceptance of an HRA method have grown significantly. Compared to method evaluations from the 80s and 90s, this study added attributes concerning accident progressions further from the design basis, organisational issues, HEP adjustments to account for uncertainties in qualitative information, and guidance concerning limiting values. One driving factor behind the inclusion of these attributes is the now widespread interest in PSA scopes beyond Level 1 and beyond internal event scenarios. Attributes were also added to reflect the importance given to the methods' technical bases, in terms of the state-of-knowledge in human factors, available data, and systematic evaluations of method reliability (consistency when applied). This importance stems in part from the increased integration of probabilistic perspectives in safety-related decision-making.

In addition to the basis in theory for a method, the human performance data underlying the development of the method and its basic HEPs is considered as an additional element of validity. Methods based on task observations and data from actual as well as simulated contexts better fulfil the requirements of this attribute, while those based on expert judgement or data pertaining to relevant tasks in other domains partially meet the attribute's requirement. With few exceptions, the intermediate rating was applied to most methods. For the situation assessment and decision component of post initiating event human interactions (Category C actions), fulfilling the requirements of the attribute will be difficult to achieve in the near term due to the challenges for data collection.

It is a feature of HRA that there has been a considerable evolution in methods since the earliest methods in the 1970s; this evolution continues today and is likely to do so for the foreseeable future. Most particularly, this evolution has been to incorporate progressively the increasing knowledge about the cognitive and decision-making aspects of operators responding to initiating or other off-normal events. As a result, the assessment of long-standing methods has tried to balance the trade-off between recognising their extensive pedigrees of application and continued acceptance versus their comparatively simple views of cognition. These trade-offs have largely been under the control of the individual reviewers and their perspectives on the relative importance of these different attributes, though the group as a whole has tried to "level" the assessments of all methods to a common perspective.

The applicability of the examined HRA methods to the analysis of accident management actions, e.g. the mitigation actions guided by Severe Accident Management Guidelines that appear primarily in Level 2 PSAs, was identified as an issue of particular interest. While the extension of several other methods could be envisioned for quantitative analysis, the only methods in the evaluation that are applicable in principle to support qualitative and quantitative analysis are MERMOS and ATHEANA. The deeper and more open-

ended qualitative analyses required by these methods can allow the analysis team to examine the relevant issues. On the other hand, this flexibility comes at the price of less structure and a quantification largely based on expert judgement. It may be that more experience with HRA in Level 2 PSA may be needed before HRA methods can provide stronger guidance for quantifying operator actions in these contexts.

The treatment of organisational factors was among the attributes considered desirable, with sub-attributes addressing safety culture and organisational process factors, respectively. Here, the ratings are based on whether an HRA method addresses these factors at all, rather than the adequacy of the treatment in an absolute sense. The positive ("high") ratings given for some methods typically indicate that they address some aspects of these factors. The issues of how to define the measures of culture and process factors as well as how to correlate these measures with performance reliability were identified as unresolved; consequently, further work and data will be needed before a consensus on whether and how to treat these factors in HRA will be reached.

Comprehensive, full-scope PSAs have increasingly become the state-of-practice, resulting in the need for HRA methods with a broader range of applicability. In light of the differences in the performance conditions associated with the various PSA scopes, the need for methods to provide guidance on appropriate limiting values on the estimated failure probabilities was identified as an evaluation attribute. In a few cases, the method's quantification model has an implicit lower limit – lower probabilities cannot be obtained. In other cases, the need to document the justification and application of a lower limit for a given HRA application scope was underscored, even if a consensus on the values appropriate for each performance context has yet to emerge.

Also in connection with the broader scope of PSAs, the information available for the qualitative analysis that underlies HRA quantification may be more limited in some HRA application scopes, relative to the basic Level 1 at-power PSA. The capability to account for uncertainties in the qualitative information was rated a highly desirable attribute. A majority of methods did not fully meet the requirements of this attribute; most of these were conceived for the Level 1 PSA. More recent methods caution the analyst to account for these uncertainties but do not propose a specific approach.

The attribute related to method reliability must be regarded as aspirational. It reflects the need to systematically evaluate the consistency and reliability of HRA method results, in terms of repeatability when performed by a given analysis team as well as by different teams. Practically none of the methods have been subject to a comprehensive reliability study; hence the predominance of "low" ratings in the evaluation reflects a lack of evidence rather than evidence of unreliability. A related issue, with a similar status, is the validation of HRA method predictions against empirical data. Here, some methods have been evaluated in the International HRA Empirical Study and its follow-up, the U.S. Empirical Study. While these studies have provided some results on empirical validity, they still represent only an initial step towards empirical validation. Weaknesses in the empirical basis or in the evidence for the empirical validity (of the results obtained with a method) may partially be compensated by independent verification and peer review. Several of the more recent methods have been subject to such review, with the findings used to revise the method. Such a review is considered to be a good practice for method development.

Inevitably in a project such as this, there are caveats and limitations to the study. The study set out to provide a pragmatic review of a set of HRA methods evaluated against a set of desirable attributes of HRA methods identified and developed by an international team of HF, HRA and PSA experts. The task group believes that this report provides useful information that can inform judgements on the selection of HRA methods for particular risk assessment applications. In the majority of cases, a consensus judgement on the way in which each HRA method addresses each attribute is achieved. Importantly the method evaluations document the basis of the agreed evaluations which provides useful information for readers to inform their own judgements of the suitability for each of the methods to address the HRA applications they wish to undertake.

1. INTRODUCTION

1.1 Purpose

This report presents the results of a joint task of the Working Groups on Risk Assessment (WGRISK) and on Human and Organisational Factors (WGHOF) of the OECD/NEA CSNI, to identify desirable attributes of Human Reliability Assessment (HRA) methods, and to evaluate a range of HRA methods used in OECD member countries against those attributes.

The purpose of this project is to provide information that will support regulators and operators of nuclear facilities when making judgements about the appropriateness of HRA methods for conducting assessments in support of Probabilistic Safety Assessments (PSA). The task was performed by an international team of Human Factors, HRA and PSA experts from a broad range of OECD member countries.

As in other reviews of HRA methods, the study did not set out to recommend or promote the use of any particular HRA method. Rather the study aims to identify the strengths and limitations of commonly used and developing methods to aid those responsible for production of HRAs in selecting appropriate tools for specific HRA applications. The study also aims to assist regulators when making judgements on the appropriateness of the application of an HRA technique within nuclear-related probabilistic safety assessments.

The report is aimed at practitioners in the field of human reliability assessment, human factors, and risk assessment more generally.

1.2 Background

The modelling and quantification of human error probabilities for use within Probabilistic Safety Analysis (PSA) is widely recognised as both an important and challenging aspect of nuclear safety assessment. To date, a large number of methods have been developed to support HRA. A report commissioned by the UK Health and Safety Executive (HSE) [1] identified over fifty extant HRA methods, thirty-five of which were considered to be potentially useful for high hazard safety assessments. In addition a number of new methods have been developed or come to greater prominence since the report's publication in 2009.

The HSE report referenced above is one of a number of published reviews of HRA methods. Some of the more notable include the Human Reliability Assessors Guide produced by the Human Factors in Reliability Group in the United Kingdom in 1988 [2] and Alan Swain's 1989 Comparative Evaluation of Methods for Human Reliability [3]. More recently the US Nuclear Regulatory Commission (USNRC) has published companion documents that identified HRA good practices [4] and reviewed commonly used HRA methods against the identified good practices [5]. A further review of a more limited number of HRA techniques has also been undertaken under the auspices of the USNRC-sponsored IDHEAS project [6]. Whilst much of the focus on these reviews has been on the applicability of HRA methods to support PSA in the nuclear industry, reviews of HRA techniques have also been conducted in other industries. For example NASA undertook a review of HRA methods to consider their ability to support safety analysis in the context of space operations [7].

The more recent USNRC-supported HRA reviews identified above were, perhaps not surprisingly, focussed on HRA methods that are commonly used in the USA in support of nuclear power plant PSA. This review has also included methods identified to be in use in several other OECD NEA member countries. The set of HRA methods subject to review is listed in Section 2.2.5 below.

In addition to the above method reviews, the HRA Empirical Studies have provided evidence on the qualitative and quantitative performance of a number of methods based on simulator data. The International HRA Empirical Study examined 13 HRA methods [8]. A follow-up study, the US HRA Empirical Study, looked at 4 of these methods [9-10]. While these studies are not validation studies, they produced empirically based evaluations of method performance, strengths and weaknesses.

1.3 State of the Art in Human Reliability Analysis

It is a feature of HRA that there has been a considerable evolution in methods since the earliest methods in the 1970s; this evolution continues today and is likely to do so for the foreseeable future. Most particularly, this evolution has progressively incorporated the increasing knowledge about the cognitive and decision-making aspects of operators responding to initiating or other off-normal events. Following the Three Mile Island accident in 1979 and its revelation about failures in cognition, the early HRA methods used various simple time/reliability correlations (T/RCs) as a pragmatic interim approach to represent the likelihoods of operator failures in decision-making. This approach became superseded in the 1990s by methods that used a few PSFs to provide an adjustment for a nominal human error probability to represent failures in decision making; the selection of these PSFs was often based on some kind of information processing model, and include procedures, stress, complexity and so on. More recently, methods have incorporated later understanding of how cognition is accomplished and, more importantly, how it can be misled by contexts. The definition of these contexts may include factors like time available but incorporate aspects of the plant and team behaviour as well as the traditional human-factors PSFs.

As a result, the assessment of long-standing methods has tried to balance the trade-off between recognizing their extensive pedigrees of application and continued acceptance versus their comparatively simple views of cognition. These tradeoffs have largely been under the control of the individual reviewers and their perspectives on the relative importance of these different attributes, though the group as a whole has tried to "level" the assessments of all methods to a common perspective as described in Sections 2.3.4 and 2.3.5 below.

1.4 Scope

The scope of the project relates solely to the HRA component of risk assessment and safety analysis. The attributes defined in the project, therefore, relate to the assessment of methods developed for quantification of human error in relation to safety actions modelled within a PSA. The project does not offer criteria relating to wider issues such as how the PSA identifies the required human-based safety actions to be assessed and the modelling of these within the PSA.

The attributes also do not differentiate between types of PSA (e.g. for different operating modes, full power, low power, shutdown), or different PSA levels, (e.g. level 1 estimates of core damage frequency, Level 2 estimation of release and Level 3 societal impacts). However we do focus predominantly on full-scope PSA requirements, as we recognise that screening applications may not meet all of the attributes as they aim to provide simplified but conservative analysis.

2. STUDY PHASES AND STEPS

2.1 Overview

The overall project was completed in two phases: Phase 1 of the project comprised the generation of a set of attributes and a methodology which could be used to evaluate HRA methods currently in use within member countries. Phase 2 of the project involved the application of the developed evaluation methodology in order to derive a consensus evaluation of the set of HRA techniques addressed within the project. Both phases of the project were undertaken via workshops led by the project Steering Group with individual method evaluations being led by small teams of reviewers. The task group members represent an international HRA community of practice, comprising nuclear industry regulators with an HRA or risk assessment technical background, HRA/PSA practitioners from utilities, technical support organisations and academics in the field.

Each method evaluation consists of the rating of the method on each of the attributes, with a justification of the rating and further commentary on how well the method satisfies the attribute or on its limitations.

2.2 Phase 1: development of attributes and evaluation method

2.2.1 Generation of attributes

The attributes were defined via two workshops, during which the individual experts identified important method characteristics and selection criteria based on their knowledge of individual country regulatory guidance, HRA practice, extant HRA method evaluations and personal judgement. The final set of twenty attributes used in the study is a consolidation and rationalisation of the outputs of those workshops agreed by all participants.

The full set of attributes is presented in detail in Section 3 of the report. It is recognised that the set of attributes is not exhaustive and that other criteria could be considered when evaluating the suitability of an HRA method for nuclear risk assessment application. However it is considered that the attributes provide a sufficient set to allow the most important issues with respect to the conduct of HRA in the context of nuclear risk assessment to be evaluated. In order to support this judgement, an exercise was undertaken to map the attributes derived in this project against the set of attributes or criteria used to evaluate HRA methods in other studies of this nature. Four studies were used as the basis for this mapping exercise:

- Towards an Improved HRA Method (Hendrickson et al 2011) [6].
- Good Practices for Implementing HRA (Kolaczkowski et al 2005) [4].
- HRA Methods Selection Guidance for NASA (Chandler et al 2006) [7].
- Comparative Evaluation of Methods for HRA (Swain 1989) [3].

The mapping exercise concluded that the list of attributes derived in this study address all of the HRA attributes or criteria used in the earlier evaluations. On the other hand, some attributes or dimensions specific to this study reflect more recent priorities in member countries. These include, for instance:

- the method's capability to treat "*Fault progressions including consequential faults and accident sequences encompassing Level 1 and Level 2 PSA*" (part of Attribute 8);

- consideration of organisational factors (Attribute 11);
- how the method deals with uncertainties related to qualitative aspects of the analysis (Attribute 17).

2.2.2 Applying the attributes

In order to determine how best the attributes could be used to evaluate HRA methods, the evaluation approaches in the above-mentioned HRA evaluations were reviewed. This review highlighted that the evaluation methods used tend to be qualitative and include either yes/no responses, or ratings on a qualitative scale supported by a discussion of how well the method meets an assessed attribute.

Workshop discussions agreed that the development of a quantitative evaluation of the methods, in other words, the definition of an overall score based on all attributes ratings would be problematic and add little value. The main challenges to the development of a scoring equation are two-fold. First, the weighting of the attributes would depend on the intended application of the HRA method and its requirements. Second, in a weighted sum, positive evaluations of some attributes could mathematically compensate for poor evaluations of other, unrelated attributes.

Consequently, the consensus was that a qualitative evaluation of the methods would provide the basis for a decision-maker to make an informed selection of methods for a specific HRA application, based on the requirements pertinent to this application.

This resulted in the development of a predominantly three-point scale (high, intermediate, low), with a two-point scale being used for a smaller number of attributes (high, low). To underscore the assigned rating graphically, the high, intermediate, and low ratings are shown in dark blue, medium blue, and light blue, respectively. In the method evaluation scale, the general interpretation of the ratings is as follows:

Rating	**Interpretation of the rating**
High	The high rating, coded in dark blue, indicates that the requirements of an attribute have been fully or largely met for the method's intended scope of application.
Intermediate	The intermediate rating, coded in medium blue, indicates that a method meets some, but not all of the requirements of the attribute.
Low	The low rating, coded in light blue, indicates that the requirements of the attribute are not met or that no evidence is available.

An important and essential component of the evaluation is provided by an accompanying narrative that explains the basis of the evaluation against the attribute.

It was recognised by the project group that some of the attributes have a wide scope that could not be addressed by a single evaluation. These attributes are broken down into sub-attributes and the evaluation is undertaken at the sub-attribute level; an example is the attribute related to the treatment of performance shaping factors (PSFs).

To support the evaluation process and to aid consistency between evaluators, scale point anchors describing the requirements for the assignment of the high, intermediate and low ratings were provided for each attribute/sub-attribute. These scale point anchors were developed by the steering group for the task and were refined based on comments raised by task group members and application of the evaluation scale during the method review phase of the task. The final attribute evaluation scales are discussed in Chapter 3; the set of evaluation forms are shown in Appendix 1.

2.2.3 Attribute grouping

It was recognised that the individual attributes derived in the workshops are inter-related to a greater or lesser extent. Therefore, to aid the evaluation process, the twenty attributes were grouped into five higher order categories as shown in Table 1 below:

- Construct Validity – a measure of the internal validity of the method which assesses the extent to which the HRA method measures or assesses what it claims to, and is consistent with an underlying theoretical model or dataset.
- Content Validity – a second measure of internal validity which assesses if the HRA method measures or assesses important determinants of human reliability.
- Empirical Validity – a measure of the extent to which numerical outputs from an HRA method have been demonstrated to correlate with other sources of human reliability data.
- Reliability – a measure of extent to which an HRA method produces consistent quantitative and qualitative output and the extent to which the derivation of the output can be traced and therefore verified by a reviewer.
- Usability – a measure of the extent to which an HRA method provides clear guidance for its application, useable outputs and the amount of resource required for its implementation.

Table 1: Attribute grouping

Construct Validity	Content Validity	Empirical Validity	Reliability	Usability
Attribute 1 Availability of information and data relating to the technical basis	Attribute 5 Qualitative assessment	Attribute 12 Empirical validity	Attribute 13 Computer models and software tools	Attribute 15 Definition of method scope
Attribute 2 The technical basis of the method (Theory)	Attribute 6 Factors influencing human reliability considered by the method.		Attribute 14 Reliability & traceability	Attribute 16 Qualitative outputs
Attribute 3 The technical basis of the method (Data)	Attribute 7 Consideration of human error dependencies			Attribute 17 Qualitative uncertainty and quantitative conservatism
Attribute 4 Internal consistency of the method	Attribute 8 Consideration of deviations and progressions in accident sequences			Attribute 18 Availability of user documentation
	Attribute 9 Consideration of cognitive error			Attribute 19 Use of limiting values
	Attribute 10 Consideration of statistical uncertainty			Attribute 20 Resources
	Attribute 11 Consideration of organisational issues			

2.2.4 Attribute importance weighting

It was also recognised that not all of the attributes are of equal 'importance'; in addition, that judgement of importance might be a function of a reviewer's technical background (e.g. HF/PSA practitioner) or current

work focus (e.g. HRA or PSA developer/regulator). In order to gauge the importance of each of the attributes, an exercise was conducted in which two groups; one comprising regulators and the other HRA/PSA users/developers (as determined by current role), separately rated each of the attributes for importance.

Four qualitative anchored importance ratings were used in the exercise as detailed below:

- "Essential" [E]: If a method rated poorly across a number of essential criteria it would not be considered fit for purpose, and should not be applied to nuclear risk assessment.
- "Highly desirable" [HD]: this was interpreted as a *strong requirement* for methods. If a method rated poorly across a number of these attributes a method would be considered of questionable validity and unlikely to be fit for purpose.
- "Desirable" [D]: this was interpreted as including criteria that add value and which support the suitability of the method for nuclear risk assessment application.
- "Indifferent" [I]: this was interpreted as criteria of relatively low importance, and which are considered to have no affect on the acceptability of the method for nuclear risk assessment application. (These attributes may however affect the choice of method for reasons other than method validity).

Following the separate discussion the groups reconvened to discuss potential convergence of opinion. The results of the exercise demonstrated a high degree of consensus between the two groups with only two attributes being rated differently; these were attributes relating to empirical validity and resources required for a method's application. The output of the importance weightings exercise is shown in Table 2. In this table, "User's ratings" refer to the perspective of potential method users, e.g. for selection of a method to be applied in a PSA.

Table 2: Attribute importance rating

Attribute		Regulators' ratings	Users' ratings
Attribute 1	Availability of information and data relating to the technical basis	E	
Attribute 2	The technical basis of the method (Theory)	E	
Attribute 3	The technical basis of the method (Data)	E	
Attribute 4	Internal consistency of the method	HD	
Attribute 5	Qualitative assessment	HD	
Attribute 6	Factors influencing human reliability considered by the method	E	
Attribute 7	Consideration of human error dependencies	E	
Attribute 8	Consideration of deviations and progressions in accident sequences	E	
Attribute 9	Consideration of cognitive error	HD	
Attribute 10	Consideration of statistical uncertainty	HD	
Attribute 11	Consideration of organisational issues	D	
Attribute 12	Empirical validity	E	D
Attribute 13	Computer models and software tools	E	
Attribute 14	Reliability & traceability	HD	
Attribute 15	Definition of method scope	HD	
Attribute 16	Qualitative outputs	HD	
Attribute 17	Qualitative uncertainty and quantitative conservatism	HD	
Attribute 18	Availability of user documentation	D	
Attribute 19	Use of limiting values	D	
Attribute 20	Resources	I	E/I

Notes:
1 Users: method users' perspective.
2. E: essential – HD: highly desirable – D: desirable – I: Indifferent.

The importance ratings are provided as additional information alongside the method evaluations which can be used by readers in determining the suitability of an HRA method for an intended application. It is not intended that the importance ratings be combined with the method evaluations in a quantitative manner; indeed it is considered that any quantitative formulation would be highly dependent on the specific HRA application for which an HRA method was being proposed.

2.2.5 Selection of HRA methods to be evaluated

Recognising the number and scope of HRA methods currently available, and the project timescale and resources available, an initial poll of task group members was undertaken to identify the HRA methods currently in use in member countries, and those for which there is interest and potential for application. This produced an initial shortlist for discussion, which included "first generation" or older HRA methods, and "second generation" or more contemporary methods. The methods selected for evaluation are presented below:

- Technique for Human Error Rate Prediction (THERP) Family – comprising THERP + Accident Sequence Evaluation Programme (ASEP) [11-13].
- Enhanced Bayesian THERP (EBT) [14-15].
- A Technique for Human Event Analysis (ATHEANA) [16-17]
- Méthode d'évaluation de la réalisation des missions opérateurs pour la sûreté (MERMOS) [18-21].
- Nuclear Action Reliability Assessment (NARA) [22-23].
- Simplified Plant Analysis Risk – Human Reliability Analysis (SPAR-H) [24-25].
- Human Cognitive Reliability (HCR)/Operator Reliability Experiments (ORE) and Cause Based Decision Tree (CBDT) Methods [26-27].
- Cognitive Reliability and Error Analysis Method (CREAM) [28].
- Failure Likelihood Index Methodology (FLIM) [29–30].
- Human Reliability Evaluator for Control Room Actions (HuRECA) [31].

Of note is that the group of methods included is unique; they have not been evaluated in a single exercise prior to this study. A short description of each of these HRA methods is included in Chapter 4 of this report.

2.3 Phase 2 conduct of HRA method evaluations

2.3.1 Input from method developers

Prior to undertaking the detailed HRA method evaluations, a principal developer of the method was contacted and invited to provide information on how the particular method addressed each attribute. Developers were not asked to rate the method against the scale or sub-scale anchor points, but rather to identify where in a method's documentation or other source (e.g. a peer review study) evidence in relation to the attribute could found. The aim of this exercise was to allow method developers to identify the best evidence available in relation to each of the methods in order that the task group could provide the most efficient and accurate evaluation of the method. In this way the information provided by developers was used as an in-feed to the task group method evaluations. The information provided by developers is not reported in full in this report. It should be noted that whilst developers for all methods, except for the THERP family methods, were contacted and offered the opportunity to contribute to the project, not all method developers were able to respond positively to the request. Information was provided by developers for the following methods:

- Enhanced Bayesian THERP.
- ATHEANA.
- NARA.
- SPAR-H.
- HCR/ORE/CBDT.

Once the evaluation of individual methods was complete, those method developers who had provided information to the project were invited to comment on the final evaluation. The aim of this step was to identify any factual inaccuracies in the evaluations. Method developer's comments on the final evaluations are presented in Appendix 3 of the report. This additional information should be taken into consideration when making decisions about the appropriateness of a method for any particular application.

2.3.2 Method evaluation process

Evaluation of the individual HRA methods selected for assessment in relation to the identified desirable attributes of HRA techniques was conducted in a number of stages. These comprised:

- Evaluation team assessment.
- Task group assessment.
- Consistency review.

2.3.3 Evaluation team assessment

An evaluation team for each method was formed, which, in almost all cases, comprised a minimum of a lead and second reviewer (who provided the peer review of the lead evaluation). Evaluation teams were selected based on their knowledge and experience of the method as well as interest. At this stage the full method evaluation scale was used, with the evaluation team providing a justification for the evaluation given for each attribute or sub-attribute. It should be noted that reviewers of each method were directed to review the method as it is described in a user manual or technical basis document. Reviewers were directed not to consider local modifications made to a method in order to improve its use, even if such modifications are common in their experience. Of note is that method developers from within the task group were excluded from this component of the evaluation for their own method.

Initial reviews were completed by the evaluation team outside of the main task group workshops. Once an evaluation team was content with their evaluation of a method these were made available for the remaining members of the task group to review via an OECD web portal and comments were provided to lead reviewers by e-mail.

2.3.4 Task Group Assessment

In order to derive a task group consensus on the evaluation of each method two workshops were undertaken in order to review the individual method evaluations. In these workshops the lead reviewer from each evaluation team presented the review of the method for consideration by the remainder of the task group with the aim that a consensus could be reached on the evaluation of each method against each attribute. For each method the lead or second reviewer (where the lead reviewer was unable to attend a workshop) provided an overview of the method and presented the evidence and justification for the evaluation in relation to each attribute. Task group members were invited to provide any additional information in relation to the attribute and the individual attribute evaluations were either agreed or changed to reflect the discussion.

Once a method evaluation had been discussed by the task group, an updated method evaluation was produced to reflect the decisions taken in the workshop. These revised evaluations were then made available via the OECD web portal for final review. Whilst the aim of the project was to achieve, for all methods, a consensus view on the evaluation of each attribute, this was not achieved in some cases. Where consensus could not be achieved, the evaluation of the method as provided by the lead reviewer was recorded on the method evaluation scale and counterarguments raised in the workshops are reported in the discussion of the report.

2.3.5 Consistency review

Once all individual method evaluations had undergone the Task Group assessment process, a summary table reporting the complete set of method evaluations was produced. This allowed for the comparison of individual method evaluations across each attribute. A copy of the final method evaluation summary table can be found in section 4 of the report. Using this summary table as an input, an exercise was undertaken to identify any apparent inconsistencies in the treatment of methods when evaluated against each attribute. This consistency review was completed as part of a workshop and was used as an opportunity to challenge any individual evaluation on the basis of a comparison against the ratings applied across the set of methods. The consistency review identified approximately 20 individual evaluations from a total of 300 evaluations where the original evaluation was challenged. Each of these challenged evaluations was discussed and a documented decision was made on whether to suggest a change to the evaluation or not. Where a change to the method evaluation was proposed the method evaluation scale, including the justification for the evaluation, was updated to reflect the decision taken at the meeting. Each updated method evaluation was then returned to the Lead Reviewer to confirm the evaluation. Three of the proposed changes were rejected by Lead Reviewers which resulted in 17 evaluations being changed as a result of the consistency review. Once the full set of methods evaluations had been updated these were circulated to the complete task group for final review.

2.3.6 Note on the evaluation of the HuRECA method

One method (HuRECA) did not follow this method evaluation process and instead was reviewed collectively the task group on the basis of a presentation of the method made by the principal method developer. This alternative evaluation process was deemed necessary as the method is a recent development and documents describing its technical background and method of application are only available in Korean. The evaluation of the HuRECA method was completed over 1 day as part of a Task Working Group meeting held in November 2012.

3. THE ATTRIBUTES

This section of the report presents the attributes derived in the study to evaluate individual HRA methods. The wording of the attributes and the scale point anchors used to direct the evaluation of a method against each attribute are discussed. As shown in Section 2.2.3, the attributes cover five broad factors that can be used to evaluate an HRA method and the individual attributes are presented within these groupings.

3.1 Construct validity

These attributes are:

- Attribute 1 Availability of information relating to the technical basis of the method.
- Attribute 2 The technical basis of the method (Theory).
- Attribute 3 The technical basis of the method (Data).
- Attribute 4 Internal consistency of the method.

They provide a measure of the internal validity of the method which assesses the extent to which the HRA method measures or assesses what it claims to and is consistent with an underlying model or dataset. This group of attributes sets out to determine the extent to which a method is underpinned by appropriate scientific bodies of knowledge and/or relevant data. It also considers the internal consistency between a method's technical basis and the qualitative and quantitative steps required to complete a HRA using the method.

Attribute 1	**Availability of information relating to the technical basis of the method**: Information is provided on the technical basis of the method, in terms of its scientific underpinnings and data, in order to allow a judgement on the validity of the method to be made.
High	Comprehensive information on the technical basis and/or data underpinning the method is available and its application is discussed as part of the documentation of the method.
Intermediate	The method provides references that allow the information forming the technical basis and/or the data underpinning it to be obtained.
Low	The method does not provide sufficient information to allow its technical basis and underpinning data to be accessed for review.
Comment	This attribute assesses the availability of reliable and complete information about the method such that users can: (1) understand whether the method has an appropriate technical basis and (2) judge that the method is appropriate to be used for their particular application. It is considered important that HRA method users should have access to material that will allow them understand the technical basis of a method in order that they properly appreciate its intended scope and method of use, and, perhaps more importantly, its limitations in relation to the human actions that they wish to assess.
Importance rating	Essential

Attribute 2	**The Technical Basis of the Method (Theory)**: The technical basis of the method is based upon, and does not contradict, a relevant body of scientific knowledge
High	The method operationalises a relevant model of human performance or system safety which has scientific acceptance.
Low	Elements of the method are inconsistent with an accepted scientific model of human performance or system safety.
Comment	This attribute relates to the underpinning (scientific) validity of the method and hence the credibility of its output. The attribute tests the extent to which a method has a theoretical basis which is consistent with accepted scientific knowledge. Whilst HRA methods are likely to be based on models of human behaviour and cognition, we also recognise that other theoretical frameworks may underpin methods either currently or in the future, and that developments should not be constrained by prescribing specific theoretical traditions. For example we recognise HRA development work using other frameworks of analysis such as complex adaptive systems would not be necessarily tied to the theories of cognitive psychology. It is also recognised that HRA method may not provide a direct operationalisation of a single theory and consider that it is acceptable for a method to draw on a range of theories to provide a technical underpinning. In addition, different HRA models have different purposes—there is no one all-embracing technique that addresses all HRA and PSA needs. Therefore the assessment of this attribute needs to be performed in light of the intended use of the particular method. This also relates to the era of development; such as the difference between the basis of base-HEP/PSF models (like THERP) versus those that adopt a holistic contextual approach like ATHEANA and MERMOS).
Importance rating	Essential

Attribute 3	**The Technical Basis of the Method (Data)**: Where the technical basis of the method is based on a dataset, the source of the data/information and its relevance for application in the nuclear industry should be demonstrated.
High	The data underlying the method are largely based on observations of actual or simulated task performance in nuclear industry tasks.
Intermediate	The data underlying the method are based on expert judgement or observations of human performance for relevant tasks in a domain that is closely related to the nuclear industry e.g. other high hazard industries.
Low	The data underlying the method are taken from tasks that are not related or relevant to nuclear industry tasks.
Comment	In this attribute, there is an issue of the degree to which the data sources are tied directly to the nuclear plant experience (and which types of plants vs. different plant applications). As a result, the definition of the high rating was expanded in discussions to add the word "…largely based on observations" in the definition of the basis for the data. No one method used exclusively nuclear plant experience but several used a predominance of plant data. A second important factor in assessing this attribute is the extent to which numeric values used in a method are based on the collation of direct observations or are derived using expert judgement. Where numeric values used in methods are based on expert judgement rather than collated observations of human performance then an intermediate rating is applied. In addition, the word "data" was understood to include all relevant sources of information and is not limited to simply statistical counts.
Importance rating	Essential

Attribute 4	**Internal Consistency of the Method**: The method demonstrates internal consistency between the technical basis, the error definition, the PSFs and the qualitative and quantitative method steps
High	The qualitative and quantitative component parts of the method are theoretically compatible and form a coherent consistent whole.
Low	There are theoretical inconsistencies between the qualitative and/or quantitative components of the method.
Comment	None.
Importance rating	Highly desirable.

3.2 Content validity

These attributes address a second form of internal validity and consider whether the HRA method measures or assesses important determinants of human reliability. Seven individual attributes are used to assess what are considered to be the most important elements of a human reliability assessment method:

- Attribute 5 Qualitative assessment.
- Attribute 6 Factors influencing human reliability considered by the method.
- Attribute 7 Consideration of human error dependency.
- Attribute 8 Consideration of deviations and progressions in accident sequences.
- Attribute 9 Consideration of cognitive error.
- Attribute 10 Consideration of statistical uncertainty.
- Attribute 11 Consideration of organisational issues.

Attributes 6, 8, and 11 are additionally broken down into sub-attributes.

Attribute 5	**Qualitative assessment**: It is recognised good practice that HRA quantification is supported by qualitative analysis to develop an understanding of operator performance within the scenario that is being assessed. This attribute considers the extent to which the qualitative analysis stages of the HRA (e.g. task analysis and error identification) is directed or prescribed by the HRA method, beyond providing a set of performance shaping factors to be considered.
High	The method contains or prescribes a process for conducting qualitative assessment**.**
Intermediate	The method includes a general statement indicating that a qualitative assessment should be provided, e.g. by referring to the use of task analysis**.**
Low	The method does not make any reference to qualitative analysis**.**
Comment	The purpose of this attribute is to assess if the method provides a means to assess qualitatively the situation being modelled in the HRA application. The attribute reflects the fact that HRA methods have different areas of focus. Some HRA methods focus predominantly on the derivation of HEPs and whilst they recognise that qualitative assessment is necessary for this, they do not provide detailed guidance on these steps of the HRA process other than identifying a set of factors that can impact the HEP. Other methods provide greater guidance on the assessment of PSFs by detailing a qualitative scheme for their assessment in order to assess their influence quantitatively. A further set of methods, that may be considered more complete HRA methods, prescribe a complete approach for undertaking task and error analysis as well as providing guidance on the qualitative assessment of the factors that are considered important in influencing the type and frequency of errors that might occur. In the course of discussions, the description of the attribute was changed from "develop an understanding of the drivers of operator performance within the scenario" to more simply "develop an understanding of operator performance within the scenario". This was to make clear that we are not specifying a set of PSFs as "drivers", which vary among the models used, but to keep the description to be theory-neutral.
Importance rating	Highly desirable.

Attribute 6	**Factors influencing human reliability considered by the method**: The method should be quantitatively sensitive to a majority of accepted factors (PSFs) that influence human reliability.
Subscale 1	**Adequacy of PSFs**. There are pre-defined lists of PSFs/EPCs available throughout the literature and within HRA methods (typically 1st generation methods). This attribute does not seek to define such a list, so as to accommodate developments in human performance, system safety and accident analysis. The evaluation teams are asked to use professional judgement when considering this attribute. It is not expected that methods will be assigned a high rating (dark blue) if only a small number of factors are accommodated, but we do not prescribe an absolute number of factors that are required.
High	The method requires qualitative assessment of a majority of accepted factors that affect human reliability (PSFs).
Low	The method does not consider a majority set of factors that affect human reliability.
Comment	Assessing this attribute involves a significant degree of judgement since the relevant factors that are important depends very much on the underlying theory of the method and the type of human failure and particular context being modelled. In addition, some methods like THERP identify very specific PSFs as factors to be modelled and others like ATHEANA depend exclusively on the situation being analysed—typically a difference between first and second-generation methods, though there is considerable variety among the first generation methods. Therefore the assessment needs to include consideration of the underlying theory and the type of application to judge whether the factors are adequate for that theory and application type.
Sub-scale 2	**Quantitative sensitivity**.
High	The method is quantitatively sensitive to the effect of each individual PSF considered qualitatively.
Intermediate	The method is not quantitatively sensitive to individual PSFs, but makes a single adjustment to the HEP based on the contribution of the overall combination of the PSFs considered.
Low	The method is not quantitatively sensitive to PSFs.
Sub-scale 3	**Interaction between factors**: Typically HRA methods adopt a linear multiplicative combination of PSFs. It is recognised that some PSFs may interact in other ways, e.g. a step change in the effect of one PSF once a threshold has been reached on a second PSF, or where the effect of the combination of two PSFs is far greater than multiplicative relationship would predict or where one PSF has a triggering effect on other PSFs in a causal chain.
High	Interactions between PSFs are accounted for on the basis of knowledge of the relationship between specific PSFs.
Intermediate	Combinations of PSF effects are accounted for using a simple linear model.
Low	Interactions between or combination of PSF effects are not considered by the method.
Comment	The interest in this attribute is the degree to which changes in modelled factors are reflected in changes in the HRA and PSA such that, if improvements are made, the PSA will more realistically reflect the effective new level of human performance. Since it is recognized that the effects of factors are rarely truly linear, a more complete model will incorporate the non-linear effects more directly. It also is recognized that some methods take a holistic perspective of assessing the scenarios being assessed. In these cases, PSFs are not assessed separately and their interactions then modelled explicitly; rather, the combinations of PSFs are assessed for the context as a whole. Thus changes in the scenario will lead to the combination of PSFs being assessed in a combined manner.
Importance rating	Essential.

Attribute 7	**Consideration of human error dependency**: Modelling should include consideration of human error dependencies or common cause failures.
High	The method provides a procedure for identifying potential sources of dependence among Human Failure Events (HFEs) and/or sub-tasks of an HFE, and provides a method to derive conditional HEPs based on the systematic assessment of these sources of dependence.
Intermediate	The method identifies potential sources of dependence, but does not provide a process for linking these sources of dependence to a quantified model for deriving conditional HEPs.
Low	The method does not address dependencies and common cause mechanisms among tasks and sub-tasks.
Comment	HRA methods should include a procedure to identify and incorporate dependencies and common-cause mechanisms when assessing failure probabilities. These dependencies can be between different people within the assessment of a particular operator action (high-level task), or between multiple human actions by the same people. Some methods may provide simple conditional probabilities between multiple people or multiple actions, and others may require the consideration of different causal mechanisms that have explicit probabilities assigned. Dependence modelling is considered one of the critical areas in HRA since the effects of dependence between multiple human failures can have very large effects on the final failure probabilities. For the so-called second-generation methods, multiple actions are assessed within the definition of the context of the analysis, and therefore dependencies are considered in the integrated quantification process.
Importance rating	Essential.

Attribute 8	**Consideration of deviations and progressions in accident sequences**: The method should provide a capability to accommodate: • Deviations from nominal accident scenarios due to: (A) Plant conditions: 1. Aleatory factors, such as sizes and locations of equipment failures and time sequences. 2. Complicating factors, such as coincident failures in control, instrumentation and support systems not normally modelled explicitly in PSA models. (B) Human failure scenarios due to crew organisational & operating practices that introduce opportunities to create new failure mechanisms. • Fault progressions including consequential faults and accident sequences encompassing Level 1 and Level 2 PSA which may involve extended time sequences and degraded operating environments should also be accommodated.
Sub-scale 1	**Deviations**
High	The method provides for the qualitative and quantitative assessment of all the types of deviations in accident scenarios.
Intermediate	The method provides for the qualitative assessment of human error during fault progressions, but does not provide for the derivation of HEPs in support of this assessment.
Low	The method does not provide a means to deal with deviations in accident scenarios
Sub-scale 2	**Fault progression**.
High	The method provides for the qualitative and quantitative assessment of human errors during fault progressions including level 1 to level 2 PSA fault progressions
Intermediate	The method provides for the qualitative assessment of human error during fault progressions, but does not provide for the derivation of HEPs in support of this assessment.

Low	The method does not provide for the qualitative and quantitative assessment of human errors during fault progressions.
Comment	The ability of methods to model more complex and extended degrees of plant damage is primarily a difference between first- and second-generation methods that can represent greater degrees of complexity and wider ranges of factors affecting performance. Of particular importance in the assessment this attribute is the support provided by the method for the qualitative analysis of such fault conditions. Many HRA methods were originally developed specifically to support level 1 PSA which addresses design base accident scenarios. More recently it has been recognized that assessment of operator actions post core damage is also required, the need for this being highlighted by the recent Fukushima accident.
Importance rating	Essential.

Attribute 9	**Consideration of cognitive error**: The method should be sensitive to the factors that influence the diagnosis and decision making component of the response to an initiating event.
High	The method estimates the probability of cognitive error based on the assessment of a set of factors that are known to affect diagnosis and decision making performance
Intermediate	The method uses a simple model such as a time reliability curve as the primary factor for estimating the probability of cognitive error**.**
Low	The method provides no way of estimating the likelihood of cognitive error**.**
Comment	The assessment of cognitive error modelling is highly judgemental as there are wide degrees in which HRA methods try to account for this class of error. In terms of HRA and PSA needs, the focus here is to assess from the point of quantification the potential risks from such errors. In some applications a simplistic approach may be adequate, and in others a fine-scaled modelling of psychological phenomena that may give rise to failure are required. No method presently considered accomplishes this fine-scale modelling, and therefore the assessment is essentially a relative ranking of the degree to which the model takes account of relevant factors.
Importance rating	Highly desirable.

Attribute 10	**Consideration of statistical uncertainty**: The method should provide for statistical uncertainty analysis of derived human error probabilities.
High	The method derives uncertainty parameters from experience (either in-plant or from relevant simulator trials).
Intermediate	The method provides generic uncertainty parameters, e.g. standardized error factors
Low	The method provides no uncertainty parameters.
Comment	None.
Importance rating	Highly desirable.

Attribute 11	**Consideration of organisational issues**: The method should consider the impact of organisational issues including safety-culture factors (attitudes and behaviours), and organisational process factors (e.g. command and control structures, conflicts of interest, communication and decision making protocols on human reliability). This attribute recognises that organisational factors have a significant impact on human performance and that these impacts can be both direct (e.g., resulting from a particular command and control structure) or more indirect, resulting from attitudinal and safety culture factors. In order to reflect these different types of organisational influence this attribute is assessed using two sub-scales.

Sub-scale 1	**Safety-culture factors** (attitudes and behaviours).
High	The method provides an adequate quantitative method to adjust HEPs based on an assessment of safety culture/safety climate**.**
Intermediate	The method provides a qualitative means to assess safety culture/safety climate, but does not include a process to modify HEPs based on the assessment**.**
Low	The method does not take into account safety culture factors**.**
Comment	The area of safety culture, while being widely discussed in terms of safety performance, is not yet maturely incorporated in HRA methods. However, the newer (typically second-generation) methods are incorporating aspects of safety culture through consideration of such issues as goal conflict and organisational tension.
Sub-scale 2	**Process factors** (e.g. command and control structures, communication and decision making protocols).
High	The method provides a quantitative method to assess process factors
Intermediate	The method provides a qualitative means to assess process factors, but does not include a process to modify HEPs based on the assessment**.**
Low	The method does not take into account process factors**.**
Comment	None.
Importance rating	Desirable.

3.3 Empirical validity

A single attribute is used to assess empirical validity which is a measure of the extent to which numerical outputs from an HRA method have been demonstrated to correlate with other sources of human reliability data. Empirical validity is sub-divided into three sub-scales.

Attribute 12	**Empirical validity**: The method should demonstrate evidence of empirical validation exercises, peer review processes or community acceptance based on application and maturity. Whilst from a scientific standpoint statistical evidence would normally be expected to demonstrate empirical validity, the attribute recognises that such scientific demonstrations are limited in the area of HRA and therefore a broader range of demonstrations of validity are assessed under this attribute.
Sub-scale 1	**Statistical evidence**
High	The HEP estimates derived by the method have been shown to demonstrate good agreement with plant and /or simulator data for comparable tasks
Intermediate	The HEP estimates derived by the method have been shown to demonstrate good agreement with HEP estimates produced by other HRA methods for the same or comparable tasks
Low	The method has failed to derive comparable HEP estimates in tests of empirical validity or has not been subject to such assessments
Comment	There are few published studies that provide a strong test of empirical validity, that is studies that test the HEP values derived by the application of HRA methods against known human error data points. Kirwan and his colleagues have performed validation against data for the methods THERP, HEART and the less known JHEDI Kirwan et al, (1997) [32]. An ISPRA study (e.g. Poucet (1989) [33]) provided a test of convergent validity, e.g. where the correlation between the HEPs produced by different HRA methods and teams is assessed. Shortcomings in the implementation of this study have been identified; consequently, the results on convergence from that study should be viewed with caution (as non-definitive results). In this context, the International HRA Empirical Study [Ref. 8] represents a major recent effort to use empirical data in evaluating the predictive performance of HRA methods. The study evaluated the performance of 13 methods by means of comparisons between human reliability analysis (HRA) predictions of crew

	performance in simulated scenarios and actual crew performance outcomes. The simulator experiments were conducted at the Organisation for Economic Co-Operation and Development (OECD) Halden Reactor Project's Human-Machine Laboratory (HAMMLAB), Halden, Norway. Organizations from ten countries, representing industry, regulators, and the research community, participated. The Empirical Study examined both qualitative and quantitative performance. Qualitative performance refers to the identification of the negative drivers of performance. Quantitative performance was also examined but the number of crew observations, inherently limited for practical reasons, yield reference HEPs that, in some cases, have very large uncertainties. A comparison of predicted vs. actual HEPs was then only possible for HFEs where multiple failures were observed and the empirical or reference HEP was consequently close to 1.0. For the remaining HFEs, the small number of observations yielded uncertainty bounds of multiple orders of magnitude for the reference HEPs. Consequently, the quantitative performance of the methods weighed more strongly the ranking of the HFEs by predicted HEP, i.e. whether the method predictions ranked the failure likelihood similarly to the difficulty rankings based on the simulator observations. In conclusion, the quantitative aspects of HRA method performance were indeed addressed by the study but only as far as the reference data supported. Nevertheless, the Empirical Study was successful in identifying a number of strengths and weaknesses for each of the HRA methods. In summary, for this attribute (Empirical Validity), the International HRA Empirical Study is not considered as providing a comprehensive evaluation of the empirical validity, even if the results on the quantitative performance of HRA methods relative to data obtained have been viewed as very useful.
Sub-scale 2	**Verification/Peer review**
High	The method has been subject to peer review by a team of recognised HRA experts, and the peer review comments have been incorporated to the development of the method
Intermediate	The method has been subject to peer review by a single, recognised HRA expert, and the comments have been incorporated to the development of the method**.**
Low	The method has not been subject to independent peer review**.**
Sub-scale 3	**Application/Maturity**
High	The method has been extensively applied, internationally, for five or more years
Intermediate	The method has been applied to a limited number of HRAs
Low	The method has not yet been applied to a HRA
Comment	This attribute assesses the extent to which a method appears to have validity by virtue of its use in a wide range of settings, whilst it is acknowledge that this is not a strong test of empirical validity, it may indicate that the method has some output validity.
Importance rating	Essential (perspective of regulators)/Desirable (perspective of users)

3.4 Reliability

A measure of extent to which an HRA method produces consistent quantitative and qualitative output and the extent to which the derivation of the output can be traced and therefore verified by a reviewer.

The relevant attributes are:

- Attribute 13 Computer Models and Software tools.
- Attribute 14 Reliability and Traceability.

The latter is broken down into 3 sub-attributes.

Attribute 13	**Computer models and software tools**: If a method incorporates the use of a computer model or software tool to analyse a human action, A QA programme should be applied to ensure quality of the design and validity of the output.
High	A relevant, recognised/ accepted international standard has been applied to the software design and verification of the computer based HRA method/tool
Intermediate	The design of the computer based HRA method/tool is based upon a documented QA process, which includes software verification**.**
Low	There is no evidence that the design of the computer based HRA method/tool is based on a structured and validated software development or QA method that includes software verification**.**
Comment	None.
Importance rating	Essential.

Attribute 14	**Reliability and traceability**: The method should provide consistent qualitative and quantitative information for comparable scenarios within analysts and between analysts for similar scenarios. The method should also provide sufficient information to facilitate tracing estimates back to input assumptions.
Sub-scale 1	**Within analyst consistency/reliability**
High	A formal comparison, amenable to statistical analysis, has been undertaken to demonstrate that the same HRA analyst provides consistent answers for analyses made at different times for the same scenario
Intermediate	An informal comparison has been undertaken, which suggests good within analyst agreement for analyses made at different times**.**
Low	There is no information available to suggest good within analyst agreement for analyses made at different times**.**
Comment	It is recognised that this attribute is somewhat aspirational and, as in the consideration of empirical validity, is dependent on the conduct of scientific assessments of the application of HRA methods. At this time these are limited in number.
Sub-scale 2	**Between analyst consistency/reliability**
High	A formal comparison, amenable to statistical analysis, has been undertaken to demonstrate that different HRA analysts provide consistent answers for the same scenario
Intermediate	An informal comparison has been undertaken, which suggests good between-analyst agreement**.**
Low	There is no information available to suggest good between-analyst agreement**.**
Comment	This criterion considers the extent to which it has been demonstrated that different analysts or different teams of analysts applying a HRA technique derive similar output. It does not assess the extent to which HRA outputs are derived by a consensus process e.g. within a team of analysts as is required during the application of some HRA techniques. Again the attribute is dependent on the conduct of scientific assessments of the application of HRA methods.
Sub-scale 3	**Traceability**
High	The method provides a procedure to ensure easy, complete traceability of the estimates of human performance in the HRA, such that an independent reviewer could trace back HEPs to relevant assumptions, models and data cited in the method**.**
Intermediate	The HRA method itself does not provide a procedure for traceability, but there is sufficient information available about the method to facilitate traceability, and enable an independent reviewer to understand what was done**.**
Low	There is insufficient information available to facilitate traceability**.**
Comment	This attribute assess the extent to which an HRA method provides tools such as worksheets which requires an analyst to document the analysis in such a way that allows an independent reviewer to confirm or challenge the outputs from application of the method.
Importance rating	Highly desirable.

3.5 Usability

A measure of the extent to which an HRA method provides clear guidance for its application, useable outputs and the amount of resource required for its implementation. Usability is assessed via the consideration of six attributes:

- Attribute 15 Definition of method scope.
- Attribute 16 Qualitative outputs.
- Attribute 17 Qualitative uncertainty and quantitative conservatism.
- Attribute 18 Availability of user documentation.
- Attribute 19 Use of limiting values.
- Attribute 20 Resources.

Attribute 15	**Definition of method scope**: The scope of the method should be clearly defined.
High	The scope of the method is clearly defined in a user manual and/or technical basis document**.**
Intermediate	The scope of the method is described vaguely and some analyst judgement is required to determine its applicability to a particular human action/error**.**
Low	The scope of the method is not defined**.**
Comment	This attribute considers the extent to which the areas of application for the method are identified or prescribed by the method developers. This is considered to be particularly important to prevent misapplications. For example, some HRA methods are targeted at specific types of human action, e.g. post initiator actions. If such methods are applied to assess other types of human action, e.g. pre-initiator actions then a detailed justification would need to be produced by the HRA developer to demonstrate why the method can be applied in this context. HRA method users should justify the choice of any HRA method applied in a safety assessment, and a clear definition of a methods scope assists users in producing such justifications.
Importance rating	Highly desirable.

Attribute 16	**Qualitative outputs**: The method should produce qualitative outputs that are useful to inform human factors and safety management improvements at the plant
High	The method generates qualitative information to inform improvements to reduce the potential for human error that is explicitly related to each of the factors that are used in the method to derive an HEP.
Intermediate	The method generates qualitative information to inform improvements to reduce the potential for human error, but this is not explicitly linked to each of the factors used in the derivation of HEPs.
Low	The method does not generate qualitative information to inform improvements to reduce the potential for human error.
Comment	None.
Importance rating	Highly desirable.

Attribute 17	**Qualitative uncertainty and quantitative conservatism**: Methods should be able to reflect uncertainties related to qualitative information via conservatisms in the quantification process.
High	The method provides a mathematical procedure for adjusting the conservatism of the HEPs derived as a function of the level of certainty in the qualitative information collected during the assessment.
Intermediate	The method provides a general caution on the need to adjust the conservatism of HEPs as a function of the level of certainty in the qualitative information collected, but does not provide a mathematical procedure for doing so.

Low	The method does not address the issue of uncertainties in qualitative information and the impact of this on derived HEPs.
Comment	This attribute assesses the extent to which an HRA method provides a mechanism or guidance for the adjustment of HEPs based on the completeness or certainty of the qualitative information which underpins the HEP. For example at the early stages of the design or modification of a system, much of the qualitative information underpinning the HRA may be based on assumptions or non-detailed "high level" information. In such cases it might be expected that conservatisms will be built into the quantitative aspect of the HRA to reflect the uncertainty in the qualitative information. As the design develops and qualitative information regarding design and operation becomes more certain then quantitative conservatisms can be relaxed.
Importance rating	Highly desirable.

Attribute 18	**Availability of user documentation**: The method should be supported by a detailed user documentation e.g., manual or instructions, which describes how the method should be applied.
High	The method contains user documentation that provides a detailed step-by-step procedure for all steps in the derivation of an HEP.
Intermediate	The method contains user documentation that provides a high level description of how it is applied to derive HEPs, but not all elements of the method are detailed as step-by-step procedures.
Low	The method provides only a high level description of its method of application and or data tables for the derivation of HEPs.
Comment	None.
Importance rating	Desirable.

Attribute 19	**Use of limiting values**: The method should provide limiting values. (Relevant Good Practice documents discuss limiting values that are used in member countries).
High	The method provides limiting values and advice on their application.
Intermediate	The method provides advice on the need to limit claims on human performance but does not provide specific limiting values.
Low	The method does not consider the use of limiting values.
Comment	This attribute considers how a method deals with the issue of unrealistically low human error probabilities and seeks to determine whether a method recognizes the potential problem and, if so, provides a mechanism for limiting the probabilities to be used. Unrealistically low HEPs might be prevented as a function the method's HEP derivation approach, or additional limiting ("cut-off") values may be provided when calculated HEPs fall below an identified minimum value.
Importance rating	Desirable.
Attribute 20	**Resources**: A comparative estimate of the resources (time, cost, utility demands, level of specialist training required, level and type of knowledge required) needed to apply the method in comparison with other HRA methods.
High	The estimated cost of and time required for applying the HRA method is less than or comparable with that of other HRA methods**.**
Low	The estimated cost of and time required for applying the HRA method is in excess of that required for application of other HRA methods**.**
Comment	This attribute provides an assessment of the typical resources required to apply an HRA method in comparison with the other HRA methods considered in this study. It is recognised that any judgement on the adequacy of a method should assess resource needs in the context of the benefit that will accrue from the cost of the utilisation of that resource. In this attribute however, only the cost aspect of the cost benefit relationship

	is assessed; it is considered that the previously described attributes provide insight into the benefits that are provided by any of the individual HRA methods. It is also recognised that a method which requires low resources for its application may not on its own be sufficient for a complete HRA analysis and that other tools would be required to be used as well of the method. (For example, SPAR-H deals only with quantification aspects and needs additional HRA work in the form of task and error analysis if human error reductions are to be achieved). Therefore when considering the choice of HRA methods to be used, the resource cost is only one factor and this needs to be evaluated in the context of the required outcomes from the HRA. In considering evaluation of the resources attribute, assessors were asked to consider a number of factors in determining the time and cost used to undertake the analysis including demands on facility and operator time, the range of experts required to undertake an assessment and the training requirements associated with the method to be applied.
Importance rating	Indifferent (perspective of regulators and some users)/Essential (perspective of some users).

4. RESULTS

4.1 Introduction

This section of the report provides the results of the evaluation of each of the HRA methods addressed in the study when compared against the identified desirable attributes of HRA. For each method a short background description is provided followed by a summary of its main strengths and limitations as identified in the review when compared to other methods. The summary of findings is supported by the full method evaluation scale shown in Appendices A2.1 – A2.10; most importantly, the full method evaluation scales (worksheets) provide the reader with the justifications for each of the assigned ratings. The results section ends with a summary table which compiles the colour coded method evaluations into a single table which can be used to compare the relative strengths and limitations of all of the methods considered in the study. It is strongly recommended that readers do not use the summary table in isolation when using this report to evaluate or select HRA methods for their own application. The authors consider the evaluation worksheets with the detailed justifications to be the primary output information on which method evaluation and selection should be based.

4.2 Technique for human error rate prediction (THERP)

4.2.1 Background

The Technique for human error rate prediction (THERP) is one of the earliest methods developed to provide estimates of human reliability, particularly in relation to nuclear power plant operations pre-initiator, fault initiating and post-initiator operator actions. It is based on two different reliability models: a model of decision-making based on the time available for decision and action following an initiating event, and a model of actions taken independent of time, such as selecting the appropriate switches and operating them correctly. The first model is represented as a time reliability curve (TRC), providing a failure probability as a function of time. The second model involves identifying human actions using task analyses and assessing the probabilities of failure based on evaluating a variety of performance shaping factors (PSFs). The effects of the PSFs and the base rates of failure probabilities are based on the judgement of the method's authors but are understood to have been largely based on their experience in observing tasks from nuclear weapons assembly tasks and tasks undertaken in non-nuclear power industries.

4.2.2 Summary

THERP has considerable strengths related to methods frequently referred to as "first generation" HRA methods. Its limitations are principally in the areas that are the focus of the "second generation"[2] methods.

2. So-called "second generation" HRA methods are generally based on the principle that human failures are mostly the result of specific plant contexts or conditions that mislead operators (or other humans) into taking incorrect actions because of particular strong cognitive effects; in other words, they do not occur at random but are induced by these contexts. In contrast, "first generation" methods represent failures largely occurring randomly but whose probability is influenced by such design features as panel layout and labelling.

The method's primary strengths are:

- The wide range of error types and PSFs addressed in the second (non-time based) model (Attribute 6.1).
- Extensive supporting documentation of the basis for the judgements of the probabilities of errors and the effects of performance shaping factors by the method's authors (Attribute 1).
- Consideration of dependency of probabilities between actions (successes and failures) (Attribute 7).
- Extensive documentation including a user's manual as well as the technical basis documentation, to support new users. (Attributes 1 and 18).
- THERP has positive statistical evidence of empirical validity (for the execution component of HFEs) (Attribute 12.1).
- The method was subjected to a formal peer review that led to modifications, subsequently published in the method's final report (Attribute 12.2).

The primary limitations are:

- Limited consideration of failures in cognition (understanding what is happening in the plant and what decisions need to be made) other than through the use of the time reliability curve (Attribute 9).
- Absence of ways to evaluate the effects of deviations in plant behaviour from the nominal accident sequence other than the effects of time (Attribute 8).
- Lack of consideration of organisational culture and process factors. (Attribute 11).
- Application of THERP as described in the user manual is more resource intensive than other 1st generation HRA methods (Attribute 20).

4.3 ASEP HRA Method

4.3.1 Background

The Accident Sequence Evaluation Program (ASEP) Human Reliability Analysis Procedure was developed by Alan Swain (the prime developer of THERP) to provide a simplified version of THERP to be used in the US NRC's Accident Sequence Evaluation Program (ASEP), performed in the 1980s and 1990s. For the most part, it is based on the most commonly used components of the THERP HRA method. The search for errors and their HEPs is divided into the pre-accident and post-accident phases, like THERP. Both use a PSF approach and use a time reliability curve for diagnostic errors. Fewer failure modes are identified, concentrating mostly on errors in using procedures. While the ASEP method is generally simpler and less comprehensive than THERP, it has added a screening process such that analysts can quickly identify the potentially more significant errors quickly (in conjunction with the PSA models). In addition, the ASEP method adds an uncertainty bounds propagation computer program and the treatment of multiple abnormal events, immediate emergency actions, and symptom-oriented procedures. These more closely matched the issues surrounding nuclear plant PSAs at the time of the method's development than did THERP.

4.3.2 Summary

ASEP has basically the same potential strengths as THERP but because it focuses on fewer human error types and fewer PSFs (mostly those centred on procedure-based actions), the comprehensiveness of THERP together with the background information justifying many of the HEPs are no longer strengths for ASEP. However ASEP does provide information not in THERP, including screening models and ways to incorporate the analysis of symptom-based procedures.

The method's primary strengths are:

- A simplified means of applying the THERP-based model to minimise the effort and make it useable by PSA systems analysts (Attribute 20).
- Additional consideration of dependency of probabilities between actions (successes and failures) (Attribute 7).
- The documentation provides a computer code to calculate propagated uncertainties (Attribute 13).

The primary limitations are:

- The provision of only a few types of human error and performance shaping factors (Attribute 6.1).
- Limited consideration of failures in cognition (Attribute 9).
- Limited consideration of organisational culture and process factors (Attribute 11).

4.4 Enhanced Bayesian THERP Method

4.4.1 Background

The Enhanced Bayesian THERP method, as its name suggests, is a modification of the Technique for Human Error Rate Prediction (THERP), to provide performance-shaping factors to adjust the HEPs from the THERP time reliability curve (TRC) for diagnostic and decision-making activities. The specific PSFs are:

- Quality and relevance of procedures.
- Quality and relevance of training.
- Quality and relevance of feedback from process (MMI).
- Mental load (stress) in the situation.
- Need for coordination and communication.

The method provides processes for both qualitative and a quantitative analyses. Several analysts perform the assessment of the PSFs independently and the assessed effects are combined using a Bayesian updating process. The method is documented in journal and conference papers and in reports provided to clients. There is no publicly available formal method report though the process seems traceable using the available documents.

4.4.2 Summary

The Enhanced Bayesian THERP method overcomes one of the criticisms of the original THERP TRC, which is that it is insensitive to PSFs other than the time available for actions. The use of Bayesian updating allows for the formal combination of different kinds of expertise from different specialists.

The method's primary strengths are:

- The use of relevant PSFs as modifiers of the original THERP TRC for diagnostic failures (Attribute 9).
- The use of the Bayesian updating process to combine judgements of different technical experts (Attribute 17).
- The method has been used to support a number of HRAs in Scandinavian countries and has been subject to review by regulatory bodies in these countries. (Attribute 12.2).

The method's primary limitations are:

- Limited formal documentation of the method, both in terms of a technical basis document and a user manual (Attributes 1 and 18).

- There is no formal process to ensure the traceability of results, though in practice it seems that the steps in the analysis can be tracked. (Attribute 14.3).
- The method does not provide guidance on the use of limiting values and it appears as though the calculation procedures could result in the production of very low HEPs. (Attribute 19).
- Limited consideration of organisational culture and process factors (Attribute 11).

4.5 A technique for human event analysis (ATHEANA)

4.5.1 Background

ATHEANA was originally developed as a method to evaluate the potential risks from so-called "errors of commission", where actions are taken by operators to interfere with operating equipment (as happened with the termination of high-pressure injection at Three Mile Island in March 1979). It has since become used in more typical HRA applications. The concept of ATHEANA, like several other second-generation HRA methods, is to examine the underlying cognitive processes by which operators make decisions to take actions and how these processes can be misled by plant and other conditions into making erroneous assessments that are then acted upon. As such, it requires quite detailed understanding of how particular accident conditions will present themselves to operators (particularly the spectrum of conditions that are often subsumed into a single nominal accident scenario in most PSAs) and how these may confound operator expectations. There are no data supplied with the method since every scenario is potentially unique, but an expert elicitation process is provided in the method's documentation.

4.5.2 Summary

Like other second-generation methods, ATHEANA can be challenging to use because it requires quite detailed understanding of plant conditions during accidents, including how they can vary for what the PSA considers a single accident condition, and it requires an understanding of individual plant training, practices and procedures. The method does not provide data since the conditions leading to each human failure tend to be unique but it does provide an expert elicitation process for quantification.

The method's primary strengths are:

- The method identifies and describes human errors whose complexity is much more realistic in comparison with actual failure events seen in accidents in that it considers a very broad range of potential influences and it aims to identify the consequences of so-called errors of commission (typical of second generation methods) (Attribute 9).
- The method provides a "search scheme" by which analysts can explore PSA scenarios to identify the kinds of combinations of plant conditions that may exist in accident deviations and fault progressions (Attribute 8).
- The method provides for organisational issues (safety culture and process factors) to be incorporated (Attribute 11).
- The method provides a formal expert elicitation process for quantification that includes consideration of qualitative and quantitative uncertainties (Attribute 17).
- ATHEANA was the subject of two peer reviews that led to improvements in the method (Attribute 12.2).

The primary limitations are:

- The traceability of the analysis process is limited (Attribute 14.3).
- The method does not address the issue of lower limits on human failure probabilities (Attribute 19).

- The method requires a higher level of expertise to apply compared with other (first generation) methods and typically requires significant resources (Attribute 20).
- Because it does not use a list of separate PSFs, it is difficult to obtain sensitivities to changes in PSF values for example. (Attribute 6.2).

4.6 MERMOS

4.6.1 Background

Electricité de France (EDF) used a multidisciplinary team comprised of reliability engineers, EOP experts, HRA analysts and behavioural scientists to develop the MERMOS method,[3] a second-generation HRA method. The method is based upon the knowledge of the dynamics of the accident and the EOPs, and puts the human factor at the centre of the system. Human actions are considered as the result of the whole operational system with multiple interactions between the components (the crew, the organisation, the EOPs and the MMI) and the process. This joint system accomplishes 3 functions (Strategy, Action and Diagnosis) in order to bring the reactor in a safe condition. The failure of one of these functions can lead to the failure of the mission (human failure event). A core concept of MERMOS is a CICA.[4] This relates to the dynamic state and the orientation of the operation system that can lead to the failure of the mission if they are inappropriate and persist in time. The identification of CICAs and other significant plant conditions are largely based on extensive simulator experience. The method does not include a database but does provide a formal data elicitation process (called RETADE) to generate the failure data.

4.6.2 Summary

MERMOS was initially developed by EDF to support PSAs for the new generation of reactor plants that used computers extensively in the main control rooms, including computer-based EOPs. It should be noted that the method continues to evolve and that this review is based on the original version of MERMOS that was available for review. Documentation for the method, including its latest developments is generally proprietary to EDF and in French, though there are papers published and available in English for the version reviewed.

The method's primary strengths are:

- The method identifies and describes human errors that are much more realistic in terms of actual failure events seen in accidents, including a wide range of potential influences and it aims to identify the consequences of so-called errors of commission (typical of second generation methods) (Attribute 9).
- The method can be applied to scenarios as accident sequences progress or multiple failures and deviations occur. (Attribute 8).
- The method provides for organisational issues (safety culture and process factors) to be incorporated (Attribute 11).
- The method provides a formal expert elicitation process for quantification that incorporates qualitative uncertainties. (Attribute 17).

The primary limitations are:

- The method requires extensive resources (Attribute 20).

3. MERMOS is used to identify several related techniques that continue to be developed
4. The CICAs refer to dynamic modes of organization within the emergency operation system (people, procedures and plant) that are basically positive but may prove negative in a very specific situation. The aim is to find such specific situations.

- The method requires a higher level of expertise to apply compared with other (first generation) methods. A single systems analyst cannot apply it alone. (Attribute 20).
- There has been limited review of the method, partly due to limitations of access to proprietary information. (Attribute 12.2).

4.7 NARA

4.7.1 Background

The Nuclear Action Reliability Assessment (NARA) method is a proprietary method developed for British Energy (now EDF Nuclear Generation Limited); it builds on the concept of the Human Error Assessment and Reduction Technique (HEART) [Ref. 33] method developed in the 1980s. Both methods use a similar approach, of identifying Generic Task Types (GTTs) that have base HEPs representing the "best" conditions. These HEPs are then modified according to an assessed set of Error Producing Conditions (EPCs). The EPCs represent a range of 18 factors similar to those called PSFs in other methods, and include such factors as unfamiliarity, time pressure and low signal-to-noise ratio. Some of the data for NARA are derived from a major UK programme called CORE-DATA to collect human performance data from different hazardous industries including the nuclear industry and thus represent a substantial body of data. Other data, particularly for the effects of EPCs, come from laboratory settings (the basis for much of HEART). Even though the method incorporates these bodies of data, there is a need to make subjective assessments about the degree to which the EPCs are in effect in any situation (and thus the degree of change in the GTT HEP); the method does, however, provide guidelines on how to make such assessments.

4.7.2 Summary

NARA is a first generation method with a substantial body of data to provide the values of the HEPs calculated using it. Many of the data are developed from a human reliability data collection programme though data from laboratory experiments on human performance are used particularly to derive EPC values. Clear directions and guidance are provided in the method's documentation but these are largely proprietary to the method's sponsor. Some materials are available from conference papers and journal articles. The method has been extensively peer reviewed under sponsorship of UK HSE/ONR.

The method's primary strengths are:

- The variety of GTTs and EPCs seem to cover most scenarios identified in current power plant PSAs (Attribute 6.1).
- The EPCs include items related to safety culture and organisational process factors (Attribute 11).
- The data underpinning the method have been developed from nuclear and other industry experience, and laboratory settings and are well documented (Attribute 3).
- The method has been the subject of several formal peer reviews (Attribute 12.2).
- The method provides guidance on and suggested limiting values. (Attribute 19).

The primary limitations are:

- The scope for the method is not clearly defined. (Attribute 15).
- The method's focus is on generation of human error probabilities and does not provide guidance for the qualitative analysis aspects of HRA such as task analysis and the identification of human errors to be modelled; it is largely based on the assumption that this will be done by other tasks in the PSA (Attribute 5).
- The method has been applied in a limited number of PSAs that are not in the public domain. (Attribute 12.3).

4.8 Standardised plant analysis risk human reliability analysis methodology (SPAR-H)

4.8.1 Background

SPAR-H was developed for the US NRC in the mid 2000s period as a means for modelling human errors in a comparatively simple manner, building on earlier HRA methods. SPAR-H is built on the common first-generation HRA assumptions that human errors can be modelled using basic HEPs that are then modified using PSFs in the traditional way. However, the method distinguishes between the types of errors known as slips vs. mistakes[5] and attempts to identify PSFs that are more consistent with their different causes. Thus it attempts to move incrementally towards the cognitive perspective underlying the second-generation methods. The PSFs considered include, for example, available time, stress/stressors, complexity, and work processes. There is a requirement for the analysts to consider combinations of PSFs not being simply linear combinations for which qualitative guidance is provided. In addition, guidance is provided for considering dependencies between errors. Versions of SPAR-H have been developed for both "at power" and low power/shutdown plant PSAs.

4.8.2 Summary

The SPAR-H method was developed as an evolutionary method that recognises the challenge of modelling the causes and effects of failures in cognitive processes. However it does not provide guidance to the PSA analysts to identify new situations resulting from such errors. Extensive guidance is provided for the users of the method and the judgements required in its application.

The method's primary strengths are:

- Detailed guidance is provided on the method in general, and particularly for quantification of multiple PSFs and dependency between failures (Attributes 1, 7 and 18).
- The method includes the consideration of organisational process factors through the PSF for work processes (Attribute 11.2).
- The method has been subjected to a well-documented peer review. (Attribute 12.2).

The primary limitations are:

- No qualitative outputs are provided to assist in the PSA modelling process or to identify risk-reduction measures. (Attribute 16).
- In cases where information on a PSF is not available to the analyst, optimistic HEP estimates may result because the SPAR-H guidance allows analysts to assume that the PSF is nominal. In other words, in the absence of positive or negative evidence, the guidance allows analysts to assume that a PSF does not negatively impact performance. (Attribute 17).

4.9 HCR/ORE & CBDT Methods

4.9.1 Background

The EPRI HRA Calculator is a software package that combines several individual HRA methods for use primarily by HRA analysts when performing PSAs for utilities, often in support of regulatory submissions to US NRC. Because of the unavailability of the EPRI Calculator software to the review team, this review has concentrated on two of the methods for assessing the likelihood of failures in cognition (detection, diagnosis and decision-making) whose documentation (report EPRI TR-100259) is available for public review. These are the Human Cognitive Reliability/Operator Reliability Experiments (HCR/ORE)

5. For further explanation of these differences and their significance, see Reason, J. (1990), *Human Error,* [34] (New York: Cambridge University Press)

method and the Cause-Based Decision Tree Method (CBDTM). The HCR/ORE method uses a normalised time-reliability correlation (T/RC) to estimate the probability of crew failure in cognition based on the ratio of the time available to decide vs. the time taken by crews in the simulator to begin to take action and is described in EPRI NP-6937 "Operator Reliability Experiments Using Power Plants Simulators" (Vol. 1-3, July 1990 and January 1991). While the method supports plants developing their own simulator data to use, the document also provides distributions based on previously performed simulator trials. The CBDTM is a set of decision trees by which the analyst assesses a set of PSFs in combinations to provide estimates of cognitive failures. Failures in execution of operator actions are assessed using the THERP method that is evaluated separately in this report.

The application of the methods is based on the guidance provided in the SHARP1 documentation, EPRI TR-101711, SHARP 1- A Revised Systematic Human Action Reliability Procedure" (T 1 and T 2, December 1992) [35]. Where the attributes related to the application process rather than the methods, this source was used as the primary basis for the assessment.

4.9.2 Summary

The EPRI HRA Calculator with the use of the HCR/ORE and CBDT methods is widely used in the USA, particularly by the industry PSA practitioners. It has been used in some international studies.

Its primary strengths are:

- The HCR/ORE component of the method derives HEPs from simulator data collected in the nuclear industry. (Attribute 3).
- It is standardised for use by the industry (for whom handbooks and training are available), and thus comparisons between results for different plants can be made directly. (Attribute 14.1).
- The methods have embedded guidance in the Calculator software for users. (Attribute 13).

Its primary limitations are:

- The use of a time ratio in the HCR/ORE component of the method does not have a strong scientific basis (Attribute 2).
- The HCR/ORE component of the method does not provide a means to account for the effects of individual PSFs (Attribute 6.2).
- Explicit consideration of organisational process factors is limited in CBDTM and absent when using the generic HCR/ORE data (Attribute 11).
- Its limited ability to evaluate the sensitivity of results to PSFs in the modelling of cognition, particularly for the HCR/ORE method. (Attribute 9).
- Qualitative outputs are not provided when using the generic HCR/ORE data. (Attribute 16).

4.10 Cognitive reliability and error analysis method (CREAM)

4.10.1 Background

CREAM was developed as an interim step towards a second-generation HRA method. CREAM is comprised of two models: a screening approach and a detailed assessment. This review describes the detailed assessment. While the method uses a range of PSFs (called Common Performance Conditions [CPCs] in CREAM), these are treated as non-independent and are expected to be assessed in an integrated manner for the context of particular actions. The CPCs include such dimensions as adequacy of organisation, working conditions and crew collaboration quality as well as adequacy of MMI and availability of procedures. Operator actions are divided into four classes of tasks: observation, interpretation, planning and execution, with specific cognitive failure functions (CFFs) being identified for each task, and different CFFs can have different consequences in terms of the PSA models. Failure

probabilities with uncertainty ranges are provided for the nominal conditions of the CPCs for each CFF; the effects of CPCs not being nominal are incorporated by multiplying the probabilities by factors provided. Additionally the method provides rules for incorporating dependencies between CPCs.

4.10.2 Summary

The method extends the range of "PSFs" typically considered in first-generation HRA methods to include organisational process factors like adequacy of organisation, working conditions and crew collaboration quality. In addition, the CPCs are not assumed to be independent and should be considered as a whole related to the context in which actions are taking place. It thus represents an evolutionary step beyond the typical first-generation approach.

The method's primary strengths are:

- The range of CPCs (PSFs) is extensive and covers organisational process issues as well as typical human factors (Attributes 6, 11.2).
- CPCs are not assumed to be independent and rules for incorporating dependencies between them are provided (Attribute 6.3).
- Multiple types of failure are identified for each step in the operators' responses to events that can have different effects in the PSA models (Attribute 8.1).

The primary limitations are:

- The sources of some data used by the method's developer are not always clear and seem to be personal judgements without them being explicitly stated so. (Attribute 1).
- The method does not consider human error dependency. (Attribute 7).
- The method has been applied in a limited number of PSAs for which publicly available information is limited. (Attribute 12.3).

4.11 Failure likelihood index method (FLIM)

4.11.1 Background

FLIM is a method that is intended to allow analysts to incorporate their judgement of how the strength of seven PSFs may influence the probability of a human error based on a comparison with human errors of known failure rates and PSF values. It is based on the success likelihood index method (SLIM, NUREG/CR-3518) [30] but calculates a failure probability rather than a success probability. In this method, the failure probabilities are derived from events with known HEPs as "calibration events"; the identification of appropriate calibration values for obtaining HEPs (data from similar events) is a critical aspect of the method. Thus the method does not provide any HEPs within itself but provides reference scales for the analyst to apply their own comparison with other events. The quantification process involves two steps: (1) assigning a relative importance of each performance-shaping factor to the overall likelihood of success for the action; this is designated the performance-shaping factor weight, and (2) estimating the degree to which each performance-shaping factor helps or hinders the operator in performance of the action; this is designated the performance-shaping factor rating. The method considers seven PSFs, including adequacy of time, procedural guidance, training and experience, and complexity. One particular challenge with using FLIM is the absence of any formal documentation of the method.

4.11.2 Summary

The method has been applied in several PSAs both in the USA and Europe, though largely by the method's developer. The method does not provide a stand alone user manual, but a step-by-step procedure

for its application is contained within NUREG/CR-6144 [29]. The need for having reference events with known failure probabilities and PSF ratings would appear to have limited its use in a wider community. However, where such data exist, the method provides a robust and easily traceable analysis anchored to real events. In addition the seven PSFs represent a range of generally accepted influences such as complexity and adequacy of time that are more comprehensive than many other methods.

The method's primary strengths are:

- The variety of PSFs represent a reasonably up-to-date range of influences modelled in first-generation methods, particularly for post-accident responses. (Attribute 6.1).
- The method considers organisational process factors. (Attribute 11.2).
- The method provides for limited consideration of fault progression to Level 2 PSA. (Attribute 8.2).

The primary limitations are:

- There is very little in the way of documentation for the technical basis of the document other than that embedded in PSA reports. (Attribute 1).
- Obtaining suitable calibration events and associated failure and PSF data is very challenging. (Attribute 3).
- The traceability of the analysis using the method is limited. (Attribute 14.3).
- The method does not provide advice on the use of limiting values for human error probabilities. (Attribute 19).
- The method requires a higher level of expertise to apply compared with other (first generation) methods. A single systems analyst cannot apply it alone. (Attribute 20).

4.12 HuRECA

4.12.1 Background

HuRECA is an HRA method developed by the Korea Atomic Energy Research Institute (KAERI) to model the reliability of human actions in using computer-based procedures in the post-accident phase of operations. It uses the THERP and ASEP methods as its underpinning. The method provides a majority of PSFs based on a literature review of HRA and ergonomics; it further represents more detailed attributes of computer-based design features such as computer-based procedures and soft controls to reflect the these features in estimating HEPs. The method provides both screening and detailed assessment methods of errors in PSA applications. It also provides guidance on using task analysis and other structured analysis tools for the qualitative assessment of actions. The quantification process estimates errors for the diagnostic and execution phases of operator actions. The diagnostic phase uses the ASEP time/reliability correlation and provides for adjustments through the use of PSFs linked by decision trees. The execution phase is modelled using step-by-step analysis of individual tasks. For both phases, extensive guidance is provided as to the assessed strength of the PSFs.

4.12.2 Summary

This method is one of two methods aimed specifically at actions taking place in computer-centred control rooms. Unlike MERMOS, it is built on an incrementally improved first-generation basis in that it decomposes actions into units of operation and then assesses probabilities on base values modified by PSFs. The range of PSFs and the use of explicit reference scales allow it to be used directly without a high degree of training in HRA. It provides guidance for all stages of the HRA, including use of qualitative tools to identify actions, screening, detailed analysis, incorporation in PSA and documentation. Documentation of the method is only available in Korean though a proprietary summary is available in English. The method has not yet been applied in any PSA.

The method's primary strengths are:

- The range of PSFs is quite comprehensive and includes specifically those associated with computer-centric control rooms including computer based operating procedures and soft controls. (Attribute 6.1).
- The method is built on accepted first-generation HRA models of cognition (the THERP and ASEP T/RCs) (Attribute 2).
- The provision of explicit anchor points for assessing PSF strengths should reduce the inter-rater variability in use. (Attribute 14.3).

The primary limitations are:

- The method does not consider organisational culture or process issues. (Attribute 11).
- The method has not yet been applied in practice. (Attribute 12.3).

4.13 Summary of results

In order to allow for comparison of the colour coded evaluations of the HRA methods reviewed in the study, a cross comparison table has been provided. It is important, however, that this cross comparison table is not used in isolation without reference to the individual method evaluation scales shown in Appendices A2.1 – A2.10. Additionally, it should be noted that the initial method evaluations were undertaken by different groups of reviewers and that a consensus view on the evaluation of a method against an attribute could not be achieved in a few cases. Where consensus was not achieved, the attribute rating and colour code represent the lead reviewer's evaluation.

As identified in the introduction to the study, there is no intention that the collected evaluations be used to identify methods that have passed or failed an arbitrary criterion, rather the intention is that the information provided by the study is used to inform the selection of methods and to identify where greater justification may be necessary for a method's selection.

The task group do not consider that judgements of a method's suitability for use be made on the basis of the colour coded evaluations alone, these evaluations should be considered in the context of the written justification of the evaluation of the attribute and the aims of the particular HRA that is being undertaken. For example it may be appropriate to select a method that uses a simple method to account for human error if a PSA is being used to identify those tasks that have the greatest contribution to risk. Once a particular diagnosis or decision is identified be risk important, i.e. via analysis of cutsets, a more in-depth analysis of that particular diagnosis can be undertaken using a method that is particularly suited to understanding factors affecting diagnosis error likelihood.

There is no intention that the colour coded evaluations be transferred into scores such that a computationally based comparative evaluation of the methods can be obtained. This study provides a qualitative evaluation of the HRA methods only.

In the analysis of the HRA methods THERP and ASEP, and HCR/ORE & CBDT were evaluated using a single method evaluation scale. The resulting analysis revealed that for some attributes different attribute ratings were applicable for (assigned to) the different parts of these linked methods. As a result of this, each of the individual methods is provided with a different row in table 3 and where different colour codes are assigned the justification for this can be found in the full method evaluation scale for that family of methods.

Table 3: Summary of HRA method evaluations

Method	1. Availability of information relating to the technical basis	2. Technical Basis of the Method (Theory)	3. Technical Basis of the Method (Data)	4. Internal Consistency	5. Qualitative Assessment	6. Factors Influencing Human Reliability Considered by the Method			7. Human Error Dependency	8. Deviations and Progressions in Accident Sequences		9. Cognitive Error	10. Statistical Uncertainty	11. Organisational Issues	
						6.1 Adequacy of PSFs	6.2 Quantitative Sensitivity	6.3 Interaction between factors		8.1 Deviations	8.2 Fault Progression			11.1 Safety-Culture Factors	11.2 Process Factors
THERP															
ASEP															
Enhanced Bayesian THERP															
ATHEANA															
MERMOS															
NARA															
SPAR-H															
HCR/ORE															
CBDT															
CREAM															
FLIM															
HURECA															

Method	12. Empirical Validity			13. Computer Models and Software tools	14. Reliability and Traceability			15. Definition of Scope	16. Qualitative Outputs	17. Qualitative Uncertainty and Quantitative Conservatism	18. Availability of User Doc-umentation	19. Use of limiting values	20. Resources
	12.1 Statistical Evidence	12.2 Verification / Peer Review	12.3 Application / Maturity		14.1 Within-Analyst Consistency / Reliability	14.2 Between-Analyst Consistency / Reliability	14.3 Traceability						
THERP				N/A									
ASEP													
Enhanced Bayesian THERP				N/A									
ATHEANA				N/A									
MERMOS				N/A									
NARA				N/A									
SPAR-H				N/A									
HCR/ORE													
CBDT													
CREAM				N/A									
FLIM				N/A									
HURECA													

Rating Scale
High
Intermediate
Low

5. DISCUSSION

5.1 Introduction

This report presents the views of a team of international experts in the fields of Human Factors, Human Reliability Analysis and Probabilistic Safety Analysis on desirable attributes of HRA methods. The report identifies a set of twenty attributes by which HRA methods can be evaluated and also presents a review of ten HRA methods against these attributes. The aim of the project was not to promote the use of any of the HRA methods reviewed or to provide a relative ranking of the suitability of the methods for conducting HRA. Rather the aim of the report is to provide readers with information on which to make an informed selection of the most appropriate method to be used for their own particular HRA application or an HRA application that they are required to review.

The detailed review sheets, presented in appendices A2.1 to A2.10, present a broad consensus within the team of experts that took part in the project. The discussion in this chapter identifies the main themes that have arisen from the project and some areas where development in the area of HRA methods is required to meet the needs of the HRA and risk assessment communities more generally. The areas where experts found it difficult to reach agreement, on how HRA methods should be judged in relation to the attributes, are also discussed. Whilst the overarching aim of the project was to arrive at a consensus in the attribute evaluations for all HRA methods, this was not always possible. Highlighting these areas of debate is an important output from this project and serves to identify areas where further research in the field of HRA is required.

The twenty attributes used in the project were grouped into five higher order categories: construct validity, content validity, empirical validity, reliability and usability. The overall results of the method reviews in relation to each of these broad categories are presented next.

5.2 Construct validity

Construct validity assesses the extent to which each of the HRA methods measures or assesses what it claims to by demonstrating consistency with an underlying theory or dataset. The four associated attributes were concerned with the overall technical basis of each of the methods reviewed. The first is the extent to which method users were able to obtain information to allow them to evaluate the theoretical and empirical foundations of the method. The second concerns the theorical basis while the third addresses the applicability of any data on which it was based to the HRA problem for which the method was to be applied. In terms of importance, the expert group rated three of the four attributes used to measure construct validity as essential and the fourth attribute, relating to the internal consistency of the method, as highly desirable. This means if a method was not found to provide evidence to satisfy the requirements of the attributes, then a user would need to provide careful argument for why this method was used for a particular HRA application.

In constructing the attributes it was recognised that some HRA methods are founded more strongly on theory e.g. CREAM, whilst others have their basis more strongly in data, e.g. NARA. It is recognised therefore that not all methods would satisfy a criterion which required a method to provide an operationalisation of a single theory; nevertheless, it was considered important that methods do not

contradict relevant theory. In considering the attribute related to data underpinning the method, the focus of the evaluation was the extent of the relevance of the data for application in the nuclear industry. Thus, methods were able to meet the requirements of the attribute where the data used in their construction had been collected in the nuclear industry context. A second factor taken into consideration was the extent to which the data used to underpin the method reflected direct observations of human performance or expert judgements derived from such observations.

The results of the reviews of HRA methods against the attributes measuring construct validity revealed that in all cases methods were found to provide evidence which allowed a high or an intermediate rating to be applied for all or some of the individual attributes.

For most of the methods a technical basis document was available which described underlying theory and/or data underpinning the method. In many cases, however, these technical basis documents were proprietary to the developer or sponsor of the method, e.g. MERMOS, NARA and HuRECA. Where a full technical basis was not available for review (Enhanced Bayesian THERP and FLIM), it was considered that sufficient information to allow the technical basis of the method to be reviewed was available in research reports and reported assessments that are available in the public domain.

Few of the method reviews identified cases where a method was found to contradict a relevant body of scientific knowledge. In the majority of cases the methods were identified to be broadly consistent with a human information processing model that considers human response to be a function of perception, decision-making and action execution, and identifies performance shaping or influencing factors that have an effect on human performance. This generic human information processing approach to HRA is consistent with what are viewed as 1st generation HRA methods which typically assign a base HEP to decision making and action components of a human failure event and then modify these base HEPs by considering relevant factors that affect performance. Whilst these methods are broadly labelled as first generation methods, it should be noted that many of the more recently developed methods reviewed in this study e.g. Enhanced Bayesian THERP, NARA and HuRECA fit with this general approach to HRA. Whilst these methods do not contradict the high-level model of human information processing we would not consider them to be a direct or detailed operationalisation of a single or specific model of human cognition or behaviour. It should be noted that the review team did not consider that a valid HRA method must be a direct operationalisation of a single model of human behaviour or cognition.

A group of methods including MERMOS, ATHEANA and CREAM, are considered to be more grounded in cognitive theory and models of human error. Typically these methods consider in greater detail how an operator may fail when completing a task, by considering factors such as the operators mental model of the task and the system he or she is interacting with. These models, typically labelled 2nd generation HRA models, consider the potential contexts of operation an operator may need to deal with and how these might interact with his mental models in order to produce errors. Thus the qualitative analysis component of such methods is typically more complex than the standard 1st generation HRA approach utilising base HEPs and PSFs.

Only the HCR/ORE method was identified to contradict a relevant body of scientific knowledge, by the initial review. It was judged that the basic assumption of the methodology, that the error probability is a function of the normalised time, was without scientific basis and can in some cases lead to the generation of low HEPs that are not credible for the situation. The review acknowledges that the CBDT method was developed to support HCR/ORE for such situations.

Turning to the data aspects of the technical basis of methods, only two methods (HCR/ORE and NARA) were identified to be based on data which come from direct observations of actual or simulated human performance in nuclear industry tasks. The HCR/ORE method is based wholly on simulator data;

however, during discussions questions were raised on the ability to generalise these data from the specific simulator used for data collection to other plant and plant conditions. Whilst some of the task group experts consider that the available evidence supports the notion that operator behaviour in simulator contexts is not representative of realistic conditions, it was acknowledged that for Type C operator actions other sources of data were unlikely to be able to generate sufficient data to support HRA method development.

In the case of NARA, a large part of the data underpinning the base HEPs used in the model is drawn from data generated in the nuclear industry. It is considered a particular strength that the Technical Basis document provides a careful linking of specific data points to HEPs. It is noted, however, that particularly for diagnostic and decision-making errors a proportion of the data is drawn from simulator exercises rather than directly observed behaviour in operating environments.

All of the HRA methods reviewed were judged to show internal consistency between the technical basis and the qualitative and quantitative components of the method.

5.3 Content validity

Content validity is a second measure of internal validity, which assesses if the HRA method measures or assesses important determinants of human reliability. Seven attributes were used to assess this dimension. Three were considered essential, three highly desirable and one desirable. The attributes considered to be essential included a general attribute which considered the completeness of the total set of factors influencing human reliability considered by the method, an attribute which considered how a method accounted for human error dependency and a final attribute which dealt with the treatment of deviations and progressions in accident sequences.

For the first of the essential items, the majority of the methods were considered to include the assessment of an adequate range of factors influencing reliability given the scope of the method and its intended use. The study did not provide a definitive list of influencing factors that should be considered but often reviewers used the USNRC list of PSFs identified in the "Good Practice for Implementing HRA" document (NUREG 1792) and the ANS/ASME PRA standard [36] as a yardstick by which to judge methods. Only two methods were considered not to include an adequate set of PSFs, HCR/ORE which uses only a time based factor to determine HEPs, and ASEP which uses a small subset of THERP factors for determining HEPs. In assessing this attribute, however, it was acknowledged that the HCR/ORE method is intended to be used in combination with CBDT and that the combination of methods considers an adequate set of PSFs. Therefore a high rating is assigned to the combination.

As well as considering whether a method contained an adequate set of factors influencing reliability, additional sub-attributes considered how the factors affecting reliability were accounted for quantitatively. The majority of methods were found to be quantitatively sensitive to the PSFs they addressed, however, the HCR/ORE method was found not to be quantitatively sensitive to PSFs in an explicit manner. Few of the methods were able to account for interactions between PSFs other than by linear combination of individual PSF weights. Those that provided non-linear combinations of PSFs tended to be those methods that would be identified as 2nd generation HRA methods e.g. MERMOS, ATHEANA and CREAM. The HCR/ORE method was found not to be quantitatively sensitive to PSFs in an explicit manner.

The second attribute identified as essential for a HRA method was a facility to model dependency between human failure events and derive conditional HEPs based on this dependency modelling. The majority of methods evaluated either contained a qualitative and quantitative model for assessing dependency-coupling mechanisms or identified qualitative dependency coupling mechanisms and identified the use of a technique external to the method itself for deriving conditional HEPs. Where a method identified an external technique for deriving conditional HEPs, this was often the THERP

dependency model. The HCR/ORE & CBDT Methods only considers dependency at a high level and refers readers to the SHARP1 framework [37] for further guidance whereas the CREAM method did not address the topic of dependency and was the only method to receive a low rating.

The third essential attribute related to a method's ability to deal with deviations and progressions in accident sequences. It was recognised by the group of experts that many HRA methods were developed to support level 1 PSA and in particular to support assessment of proceduralised operator tasks either pre or post an initiating event. More recently however, it has become recognised that the demands placed on operators may be greater than those considered in traditional level 1 PSA approaches. This will include situations where an initial event may be complicated for example by a loss of instrumentation or by errors of commission which can exacerbate the event. It is also recognised that once fault sequences proceed beyond core damage, operator actions over extended time periods and in degraded operating environments may be required to prevent release of fission products. Whilst the need for HRA methods to address such scenarios has been recognised for some time, the Fukushima accident has increased the prominence of this need for the nuclear risk assessment community.

The study identified that only two of the methods reviewed, MERMOS and ATHEANA, provided adequate support in principle for the qualitative and quantitative assessment of such accident sequences. It should be noted that in evaluating this attribute a strong emphasis was placed on the method's ability to support the qualitative analysis aspects of the assessment in terms of identifying important errors that might be made due to the particular contexts in which operators would find themselves. Limited support was found for CREAM for its use in relation to some deviations in accident sequences due in main to the general applicability of the concept of common performance conditions (CPCs) although these were not considered to be adequate for progressions in fault sequences. Similarly it was considered that the qualitative guidance contained within FLIM had potential usefulness for modelling some aspects of fault progressions, however, in neither case was it considered that the technique provided a sufficient basis for the quantification of human error in these accident sequences.

Whilst it was considered that some of the PSFs contained within other HRA methods are likely to be important, there is little evidence available to indicate that the impacts of the PSFs as modelled in existing methods (e.g. multipliers) are applicable in these types of accident sequences. More importantly it is not clear that the HF, HRA and risk assessment communities have sufficient knowledge of the range of factors and the strength of their impacts in severe emergencies. An OECD NEA WGHOF project is currently underway to address this issue which should produce outputs useful to the HRA community in relation to this issue.

The three highly desirable attributes addressed the issues of qualitative assessment, statistical uncertainty and cognitive error. The attribute related to qualitative assessment considered the extent to which the HRA method provided guidance on the conduct of the qualitative analysis that is necessary to underpin quantification of human error probabilities. A high rating for this attribute would require that a method contained an explicit qualitative assessment process that went beyond the provision of a list of influencing or performance shaping factors that should be considered. The expectation was that the methods assigned a high rating would prescribe the required process for task analysis and error identification rather than simply refer to the need for these activities to be undertaken using an unspecified method.

Six of the methods reviewed in the study were considered to meet the requirement of the high rating: THERP, ASEP, ATHEANA, MERMOS, CREAM and FLIM. All of the other methods identified that qualitative assessment was required to support application of the method, but these methods only provided detailed procedures for the quantitative analysis component. The task group recognise the importance of the qualitative analysis phases of the HRA process, the fact that this attribute is rated highly desirable

indicates that whilst it is preferable that an HRA method provides a complete HRA approach we recognise that it is possible to produce an adequate HRA by integrating the output from a number of methods. In future, a project that seeks to evaluate qualitative HRA methods may usefully complement the output from this project.

Assessment of the methods in relation to the attribute related to statistical uncertainty revealed that all of the methods included a process for deriving statistical uncertainty parameters for derived human error probabilities. These typically were based on projected statistical distributions, but for two methods, ATHEANA and HCR/ORE, were based on collected data.

The final highly desirable attribute assessing content validity concerned the topic of cognitive error. This attribute assessed whether and how the method dealt with the diagnosis and decision-making component of the response to an initiating event. All of the methods included in the review were considered to provide some facility for dealing with cognitive error and as a result none of the methods received a low rating on this attribute. Method reviewers also considered whether the probability of cognitive error was assessed only on the basis of a simple model, for example a time reliability curve, or whether the method considered a set of factors known to affect diagnosis and decision-making performance. Note this attribute deals only with the quantification of error and does not assess the ability of the method to identify the different types of cognitive error that might occur, e.g. cognitive errors of commission, the issue of qualitative assessment was dealt with separately via attribute 5.

Three of the methods, THERP, ASEP and HCR/ORE were rated intermediate on this attribute indicating that method reviewers considered the method used a simple model for deriving HEPs related to cognitive error. In the cases of THERP, ASEP and HCR/ORE the basis for this decision was that the HEPs were derived on the basis of a time reliability curve and did not take into account other factors that might affect the probability of failure to diagnose.

All of the other HRA methods reviewed in the study were assigned a high evaluation for this attribute; it is recognised however, that a number of alternative approaches to the treatment of cognitive error are provided by the methods reviewed. A group of methods assigned a high rating (dark blue), e.g. Enhanced Bayesian THERP, NARA, SPAR-H, CBDT, FLIM, HuRECA and CREAM, adopt a base HEP adjusted by consideration of PSFs approach. These methods are particularly useful for quantifying the likelihood of errors of omission during decision–making and diagnosis. There was considerable debate amongst the group of experts undertaking method reviews in relation to this attribute. Some experts considered that methods of the type outlined above do not provide for an adequate set of factors related to cognition and would have preferred an intermediate rating to be applied to this group of methods.

Other HRA methods e.g. MERMOS and ATHEANA provide for a more detailed consideration of situational context in deriving HEPs related to cognitive error and provide a means by which errors of omission and errors of commission can be quantified. On the other hand, these latter methods involve a greater degree of expert judgement in the identification of HFE-specific failure scenarios and in their quantification.

The debate held within the group in relation to what constitutes an adequate treatment of cognitive error to some extent reflects the different background of the experts making up the task group. It is not surprising that those with a background in Human Factors, particularly Psychologists in this group would wish to see a more complete model of cognition to be used as the basis for the treatment of these types of error in HRA.

On other hand the degree of fidelity of the model of human cognition that is required must be balanced against the reason why the HRA is being undertaken. Any HRA and safety analysis more

generally is an iterative process which is typically undertaken in increasingly narrow and deeper slices as the analysis progresses. Initially a screening analysis may be undertaken to identify those human actions which have some appreciable impact on risk. Once this subset of human actions is identified a more detailed assessment of these actions will be undertaken, but even at this stage, a large number of actions may require to be considered. At this stage screening values will be replaced with more accurate HEP estimates and this is the type of assessment typically undertaken by 1st generation HRA methods which can provide an approximation of error likelihood based on a limited range of performance shaping factors known to affect human reliability. Once these data are entered into the PSA, cutset and importance analysis can identify those particular human actions or failure events that have the largest impact on risk and these actions can then be subject to even more fine grained analysis, perhaps using second generation HRA methods which provide for a more complete analysis of the contextual factors that can impact performance on these more risk important task. This iterative approach to HRA and safety analysis allows for a proportionate use of HF and HRA resources in the conduct of safety analysis.

The final attribute considered under the heading of content validity related to the treatment of organisational factors. The attribute dealt with two aspects of organisational factors, safety culture and organisational process factors such as command and control structures, communication and decision-making protocols, etc. This attribute was rated as desirable during the attribute development phase of the project. Only three methods were judged to allow for HEPs to be adjusted reflecting the influence of safety culture. Two of these, MERMOS and ATHEANA account for safety culture in their construction of the context which affects human performance, in MERMOS safety culture can be accounted for in the development of CICAs whereas in ATHEANA safety culture can be identified as part of the Error Forcing Context (EFC). NARA was also identified as providing some limited consideration of safety culture within the set of Error Producing Conditions (EPCs) used by the technique to adjust base human error probabilities, e.g. incentive to use more dangerous procedures, low workforce morale and adverse organisational environment.

None of the other methods reviewed in the study were judged to address the issue of safety culture. It is noted that assessment of safety culture was another area in which considerable debate was held within the task group. Some experts expressed the view that no HRA method provides an adequate consideration of safety culture and therefore considered that all methods should be assigned a low rating on this sub-attribute. Other believed that whilst the treatment of safety culture within HRA methods is a simplification of the relationship between organisational culture and human performance, they nevertheless provide for a basis from which to model the impact of some aspects of safety culture within HRA.

In comparison to the consideration of safety culture, a larger number of the HRA methods reviewed were considered to provide a capability to incorporate organisational process factors in an HRA. A significant number of methods, however, were not considered to address organisational process factors; these were ASEP, Enhanced Bayesian THERP, HCR/ORE, CBDT and HuRECA, whilst THERP was considered to provide a mainly qualitative discussion of organisational factors.

Generally for the topic of organisational factors, where a method receives a high rating this reflects the fact the some of the factors underpinning these topics are addressed within the HRA method rather than reflecting that the entirety of the factors that could be labelled as organisational are addressed.

5.4 Empirical Validity

This attribute considered the extent to which the numerical outputs from an HRA method have been demonstrated to correlate with other sources of human reliability data. In addition to considering scientific experiments to assess empirical validity the attribute also consider whether evidence of validity had been demonstrated from the conduct of peer review exercises or via risk assessment community acceptance, based

on application and maturity. This attribute was one of only two attributes that were considered to have different levels of importance by regulators and the users of HRA methods. Regulators viewed empirical validity to be essential for an HRA method, whilst users considered empirical validity to be desirable.

There are few scientific articles that report validation studies for HRA methods against either known human error data points or between HEP estimates produced by different HRA methods, so called convergent validity studies. As a result, only one method, THERP, was consider to have evidence of true empirical validity (for the modeling of execution/implementation and not for the diagnosis model). Several methods evaluated in this study were used in the International HRA Empirical Study, which represents a major recent effort at validating HRA methods. However, as discussed in the presentation of Attribute 12 "Empirical Validity" in section 3 of this report, this study is not considered a comprehensive evaluation of empirical validity, even though its use of empirical data as evidence provided useful insights of the methods' potential empirical validity. At the same time, the challenges associated with producing quantitative reference data that are adequate for empirical validation should not be underestimated. This holds in particular for such data for the decision-related aspects of human performance.

Given the limited number of studies assessing empirical validity two further sub-attributes which may reflect on a method's validity were assessed under this attribute. The first of these considered whether a method had been subject to peer review during its development. Five HRA methods, THERP, Enhanced Bayesian THERP, ATHEANA, NARA and SPAR-H, were considered to have been subject to peer review by one or more teams of recognized experts, whilst MERMOS was subject to a peer review by a single expert during its development. The final sub-attribute under this heading considered the extent to which a method's validity could be implied from a history of application in multiple settings. On this sub-attribute all of the HRA methods accept for HuRECA demonstrated some evidence of repeated use with particularly strong evidence of use being recorded for THERP, ASEP, SPAR-H, HCR/ORE & CBDT and FLIM. It is recognized of course that repeated use of a technique does not provide a true measure of empirical validity and can at best be treated as a measure of community acceptance that the method provides useful outputs.

Given that regulators view empirical validity to be an essential attribute for an HRA method and the lack of scientific studies conducted in this area this reinforces the need for further studies, such as that conduct by Kirwan et al [Ref. 22] to be undertaken within the scientific community.

5.5 Reliability

Attributes in this category measured the extent to which an HRA method produces consistent quantitative and qualitative output and the extent to which the derivation of the output can be traced and therefore verified by a reviewer. Two attributes were considered within this category, the first considered reliability of computer models and software packages used to undertake HRA, based on consideration of the standards or QA plan that supported software development. This attribute was rated as essential by all members of the task group. Only four of the HRA methods assessed in this study were supported by a software package, ASEP, HCR/ORE & CBDT and HuRECA and in all cases these were developed using a documented QA process.

The second attribute testing reliability of HRA methods used three sub-attributes measuring within-analyst reliability, between-analyst reliability and traceability of output. This attribute was rated as highly desirable by task group members during the attribute development process. Only a limited number of formal and informal studies have considered within- or between-analyst reliability, despite the general concern with these issues. For the HCR/ORE & CBDT methods, it is reported that there is some evidence of within-analyst consistency. These reports are based on informal comparisons of analyses conducted at similar plant at different times, which have shown good agreement, but are not published.

In contrast to the International HRA Empirical Study, the follow-up U.S. HRA Empirical Study did address between-analyst reliability. The study design was based on multiple teams applying each of the

four methods. The limited number of teams applying a given method (in most cases, two teams) and the small number of HFEs analyzed by the teams have the consequence that the results are indicative rather than definitive. On the other hand, the consistency among the analysis teams was examined at a high level of detail. The consistency evaluation examined both the obtained, overall HEPs, the intermediate quantitative results, and the qualitative results. Furthermore, the study examined the consistency of the assumptions made in the modelling of the HFEs.

The final sub-attribute for this attribute assessed traceability which provides a measure of the ease with which a reviewer can trace the process by which an HEP has been derived. All of the methods assessed in the study were considered to provide either a process for or sufficient information to allow an independent reviewer to trace the derivation of HEPs.

5.6 Usability

The final group of attributes developed in the study addressed a method's usability. Six attributes were constructed under this heading. Three attributes, definition of method scope, qualitative outputs and how uncertainties in qualitative information should be dealt with in the quantification process, were rated as highly desirable attributes by the task group. Two further attributes, user documentation and advice on limiting values, were rated as desirable whilst a final attribute, related to resources, was rated an essential consideration by some users but regulators were indifferent or insensitive to this measure.

The reviews of the methods identified that the majority of the methods provided detailed user documentation, often in the form of a user manual and that within these documents the methods scope was clearly identified. None of the methods reviewed received a low rating for either of these attributes and only Enhanced Bayesian THERP and NARA were assigned an intermediate rating for the availability of user documentation and definition of the method's scope respectively. All of the methods, except for HCR/ORE were considered to provide useful qualitative outputs that could inform improvements on site, although differences exist between methods in terms of the specificity of the qualitative information generated to inform the improvements. The methods THERP, ASEP, ATHEANA, MERMOS, NARA, CREAM and HuRECA are consider to provide qualitative information for improvements that is specifically linked to each of the factors used in the derivation of the HEP.

The reviews of the methods against the attribute related to limiting values identifies three methods that do not consider the topic of limiting values in their user documentation, these are Enhanced Bayesian THERP, ATHEANA and FLIM, although it is recognised that advice on this issue is available in other good practice documents such as NUREG 1792 that HRA analysts may be aware of. Other HRA methods reviewed either provide specific limiting values or have calculation procedures which effectively limit the HEP that can be generated to a value consistent with or above recommended limiting values.

The attribute which considers how uncertainties in qualitative information should be dealt with in the quantification process was developed to recognise the fact that HRA can be conducted at different stages of a facility's development. Early in the design of a plant or a modification to a plant, it is likely to be the case that detailed design information is not available to support a full HRA and that assumptions may need to be made about factors that will influence human reliability. In such cases it is considered appropriate to increase the conservatism of the assessment due to uncertainty in the information on which the HRA is based. Only a small number of the HRA methods reviewed in the project were considered to address this issue: Enhanced Bayesian THERP, ATHEANA, MERMOS and NARA. ATHEANA's elicitation process used in quantification explicitly addresses uncertainties associated with the qualitative analysis. None of the others provide a clearly articulated mathematical procedure for adjusting the HEP based on the level of uncertainty associated with the qualitative information available and thus at best achieve an intermediate rating on this attribute.

The final attribute under the heading of usability considered the amount of resource required to undertake an assessment using a method. A number of factors including, time required to apply the HRA method, access to plant staff and facilities, numbers of different experts and amount of training were considered in assessing required resources but these were then combined to produce a single relative rating of the resources needed to undertake the analysis. Three of the HRA methods reviewed were rated as requiring relatively more resources for their application than other HRA methods; these were THERP, ATHEANA and MERMOS. It should be noted that judgement of resources required does not take any account of the relative benefits achieved from the use of any method, thus it may well be the case that those methods that require comparatively more resource provide the analyst with a greater amount of useful information than other methods requiring less resource for their application. This study does not conduct such a cost-benefit analysis, although readers can use the information collected in relation to the other attributes to form a judgement in relation to the benefits associated with the application of any of the methods reviewed.

5.7 Limitations of the study

This study set out to provide a pragmatic review of a set of HRA methods evaluated against a set of desirable attributes of HRA methods identified and developed by an international team of HF, HRA and PSA experts. The study does not claim to provide a rigorous scientific analysis of the methods and a number of acknowledged limitations with respect to the study's method prevent it from doing so.

The first limitation to note concerns the organisation of the method reviews. Each method considered within the study was evaluated by a different team of method reviewers. Whilst the reviews were conducted using the same method evaluation scale, with anchor points used to support the evaluation of each attribute, it is inevitable that there will be some variability in the way in which the attributes are assessed by different reviewers. A more rigorous experimental design would have required each method to be reviewed by the same teams of reviewers, so that any biases or modes of interpretation displayed by single reviewers or review teams would have an equal effect across all of the methods. However due to the limited resource available from each task group member, a study design displaying this level of rigour was not feasible.

A second limitation of the study related to the composition of the task group at different task group meetings. The study comprised four task group meetings which were held over a two year period. One of the results of this was that task group membership changed over time. This resulted, for example, in some task group members taking part in the attribute development phase of the project but not being available to support the method reviews. Similarly some task group members were only able to attend one of the two task group review meetings which could have affected the consistency in the way in which the attributes were applied due in part to individual interpretations of the attributes and also different group dynamics that will arise when groups are formed from different individuals. Ideally, the task as a whole, or at least the task group review meetings would have been undertaken as a single session with the same group of people acting as attribute developers and method reviewers throughout.

A third limitation of the study arose from the limited time that task group members could devote to the task. The project could be considered to be ambitious in terms of the number of methods and the amount of material that needed to be reviewed. Reviews for many of the individual methods required the assimilation of large volumes of background information, such as technical basis documents, user manuals, and results of other review studies. This amount of materials made it impossible for all task group members to develop the same level of knowledge about these methods. On the other hand, other methods reviewed have been recently developed or have a narrow domain of application in terms of the countries in which the method is used and thus there was limited material to review to gain an understanding of them. Complete documentation for two methods was not available in English, the working language of the Group; these were HuRECA and MERMOS.

It was inevitable therefore that the task group was better equipped to challenge the lead reviewers for some methods compared to others. A process to increase the level of scientific rigour by ensuring that all task group members were able to read the same material about every method would have been of benefit; this was not realistic given the resource available to undertake the work.

Despite these limitations to scientific rigour, however, the task group believes that this report provides useful information that can inform judgements on the selection of HRA methods for particular risk assessment applications. In the majority of cases a consensus judgement on the way in which each HRA method addresses each attribute is achieved. Importantly the method evaluations document the basis of the agreed evaluations which provides useful information for readers to inform their own judgements of the suitability for each of the methods to address the HRA applications they wish to undertake.

5.8 Areas for further research

This study has identified a number of attributes where current HRA methods address the attribute in a limited or partial way. This provides good evidence for where further research would be appropriate to advance knowledge in relation to HRA methods. In common with other studies that have reviewed HRA methods this study has found that the scientific evidence available concerning the empirical validity and reliability of HRA methods is quite limited. This has been a known problem in the HRA community for many years and recent attempts to address this issue, e.g. the International and US HRA empirical studies [8,9] have illustrated the difficulty in trying to undertake studies to provide such evidence. Perhaps we, as an HRA community, should accept that the high quality scientific evidence needed to demonstrate true empirical validity and reliability for HRA methods is unlikely to be provided and accept other weaker forms of evidence for making judgements of validity and reliability.

A second area where the results from the study identify a need for further research is in relation to how best safety culture factors can be addressed by HRA methods. This is an area where opinion was divided amongst the task group where some members considered that some aspects of safety culture were addressed by current HRA methods and others believed that safety culture was not and perhaps could not be addressed by current or future HRA formulations.

A third area for research identified by the attributes generated in this study is the issue of how best to account for uncertainties in qualitative information. This attribute, we believe, has not been considered in previous reviews of HRA methods and this review has revealed that current HRA methods are not particularly sensitive to this issue. A small study which collected data from current HRA practitioners on how they deal with this issue when conducting HRA particularly at the design stage of plant or modifications may be appropriate.

The final attribute where the need for HRA method development was identified was in relation to deviations and progressions in accident sequences. Whilst two second generation methods were consider to be appropriate for the qualitative and quantitative assessment of human actions in such situations, there remained a view that more research was required to properly appreciate the range of factors that become important in more severe accident conditions and also the size of the impact of these factors on human errors of different types. A number of studies are underway including a WGHOF project, stimulated by the Fukushima accident, to try and gain additional insight into human performance in severe accidents.

6. CONCLUSION

The work undertaken in this study has derived a set of attributes that can be used to evaluate HRA Methods in order to aid in the selection of such methods for different HRA applications. The study was undertaken by a team of recognised experts in the fields of Human Reliability Analysis, Human Factors and Safety Analysis representing OECD member countries to enable a broad perspective of views on desirable attributes of HRA to be collated. This is considered to represent a particular strength of this piece of work.

As well as identifying the desirable attributes, the study has derived an attribute evaluation scale, with defined anchor points, which could be used by readers to undertake their own evaluations. The attributes and the attribute evaluation scale have been applied to a set of HRA methods that were identified by task group members as being used in the member countries they represent. This application of the attributes has served both to refine the attribute evaluation scale and also to provide a set of data on those methods included in the evaluation that can be used by readers to support decision-making in relation to the selection of HRA techniques. Thus, this study has two main outputs: a method for undertaking the evaluation of HRA methods against a set of identified desirable attributes; and a set of method evaluations that can be used by readers in judging the suitability of a method for an HRA application they wish to undertake.

The HRA method evaluations were conducted by small teams of HF, HRA and Risk Analysis experts and each of these evaluations was further reviewed by the task group as a whole and in the majority of cases a consensus agreement on the evaluation of a method was achieved. Instances where such consensus could not be achieved are clearly identified in the discussion of the study's findings. The production of a transparent evaluation scale allows for a readers to conduct their own evaluations both of the methods evaluated by the task group and also for other methods which were not evaluated in the study. The reporting of the method in detail also allows for method evaluations to be updated in relation to the attributes as further knowledge in relation to HRA becomes available and greater experience in using new HRA methods is established.

The study did not set out to score HRA methods or provide a direct comparison between methods in order to promote or rule out the use of particular methods for particular HRA applications. The aim of the study was to provide a method and information that could inform HRA users when selecting methods. A three-point evaluation scale was developed in the study; the rating and associated colour coding was used to highlight where more careful consideration might be required in selecting a particular method for a particular purpose.

The results of the study revealed that HRA methods generally demonstrated good construct validity by demonstrating consistency with bodies of scientific knowledge. Generally, where HRA methods are based on data, these tend to be derived from expert judgements; only a small number of techniques were found to be based on direct observations of human performance and where this was the case these were often based on the observation of behaviour in simulators rather than real operating environments.

The method evaluations identified a number of areas of content validity where HRA methods may require further development; these concern accounting for organisational issues, particularly safety culture,

and the factors that influence human behaviour during more complex deviations from expected accident sequences or in severe accident conditions.

The study has found that there is little statistical evidence in relation to empirical validity and reliability for any HRA method. For empirical validity, the scarcity of data is problematic. For reliability, it is due to the lack of comprehensive scientific studies conducted to date. Other evidence that might be used to infer validity and reliability e.g. evidence of peer review, wide application or traceability, was found to be in place for the majority of methods.

The HRA methods reviewed were generally found to be well supported by user documents which defined the scope of the methods and described their method of application in sufficient detail. An estimate of the resources required to apply a method is provided, but a caution is raised that a consideration of resource requirements must take into account why the HRA method is being used and its strengths and limitations in relation to this area of application. Information on the strengths and limitations come from the consideration of the remaining attributes identified in the study.

The report acknowledges that this study does not meet the criteria of scientific evidence due to aspects of the methodology adopted. It is considered, however, to provide a useful pragmatic review of a number of HRA methods that can be used by HRA, HF and risk assessment communities.

7. REFERENCES

1. Health and Safety Executive (2009), *Review of human reliability assessment methods*, RR 679 Research Report, prepared by the Health and Safety Laboratory, United Kingdom.
2. Humphreys, P. (Ed.) (1988), *Human Reliability Assessors Guide: A Report by the Human Factors in Reliability Group,* SRD Association.
3. Swain, A.D. (1989), *Comparative evaluation of methods of human reliability analysis*, Gesellschaft fur Reaktorsicherheit m.b.H. (GRS), Germany.
4. Kolaczkowski, A., J. Forester, E. Lois and S. Cooper (2005), *Good Practices for Implementing Human Reliability Analysis*, NUREG-1792, U.S. Nuclear Regulatory Commission, Washington, DC, United States.
5. Forester, J., A. Kolaczkowski, E. Lois and D. Kelly (2006), *Evaluation of Human Reliability Analysis Methods Against Good Practices*, NUREG-1842, U.S. Nuclear Regulatory Commission, Washington, DC, United States.
6. Hendrickson, S.M.L., G.W. Parry, J.A. Forester, V.N. Dang, A.M. Whaley, S. Lewis, E. Lois and J. Xing (2012), Towards an Improved HRA Method, *Proceedings of the 11th International Conference on Probabilistic Safety Assessment and Management (PSAM11/ESREL2012), 25-29 June 2012, Helsinki, Finland*.
7. Chandler, F.T., Y.H.J. Chang, A. Mosleh, J.L. Marble, R.L. Boring, D.I. Gertman, (2006), Human Reliability Analysis Methods: Selection Guidance for NASA, NASA/OSMA Technical Report, Washington, DC, United States.
8. Forester, J., V.N. Dang, A. Bye, E. Lois, S. Massaiu, H. Broberg, P.Ø. Braarud, R. Boring, I. Männistö, H. Liao, J. Julius, G. Parry and P. Nelson (2014), *The International HRA Empirical Study: Lessons Learned from Comparing HRA Methods Predictions to HAMMLAB Simulator Data*, NUREG-2127, U.S. Nuclear Regulatory Commission, Washington, DC, United States. (Also refer to NUREG/IA-0216.)
9. Forester, J., H. Liao et al., The US HRA Empirical Study – Draft Report – Assessment of HRA Method Predictions against Operating Crew Performance on a US Nuclear Power Plant Simulator. To be published as NUREG-2156, U.S. Nuclear Regulatory Commission, Washington, DC,United States.
10. Bye, A, V.N. Dang, J. Forester, M. Hildebrandt, J. Marble, H. Liao and E. Lois, *Overview and Preliminary Results of the US Empirical HRA Study*, Proceedings of the 11th Probabilistic Safety Assessment and Management/European Safety and Reliability 2012 (PSAM11/ESREL2012), 25-29 June 2012, Helsinki, Finland, CD/ROM.
11. Swain, A.D. and H.E. Guttmann (1983), *Handbook of Human Reliability Analysis with Emphasis on Nuclear Power Plant Applications – Final Report*, NUREG/CR-1278, U.S. Nuclear Regulatory Commission, Washington, D.C., United States.
12. Bell, B.J. and A.D. Swain (1983), *A Procedure for Conducting a Human Reliability Analysis for Nuclear Power Plants* – Final Report, NUREG/CR-2254, U.S. Nuclear Regulatory Commission, Washington, DC, United States.

13. Swain, A.D. (1987), *Accident Sequence Evaluation Program (ASEP) Human Reliability Analysis Procedure*, NUREG/CR-4772, U.S. Nuclear Regulatory Commission, Washington, DC, United States.

14. Pyy, P. and R. Himanen (1996), A Praxis Oriented Approach for Plant Specific Human Reliability Analysis – Finnish Experience from Olkiluoto NPP, *Proceedings of the Probabilistic Safety and Management'96 (ESREL'96 – PSAM-III),* 24-28 June 1994, Crete (Greece), pp. 882-887, Springer-Verlag London Ltd.

15. Holmberg, J. and P. Pyy (2000), *An Expert Judgement Based Method for Human Reliability Analysis of Forsmark 1 and II Probabilistic Safety Assessment*, In: Kondo, S. & Furuta, K. (eds.), Proceedings of the 5th International Conference on Probabilistic Safety Analysis and Management (PSAM5), Osaka, Japan, 27 November-1 December 2000, Vol. 2/4, Universal Academy Press, Tokyo, pp. 797-802.

16. United States Nuclear Regulatory Commission (2000), *Technical Basis and Implementation Guidelines for a Technique for Human Event Analysis (ATHEANA)* – Revision 1, NUREG-1624, Washington, DC, United States.

17. Forester, J., A. Kolaczkowski, S. Cooper, D. Bley and E. Lois (2008), *ATHEANA User's Guide: Final Report*, NUREG-1880, U.S. Nuclear Regulatory Commission, Washington, DC, United States.

18. Blirando, C. and H. Pesme (2002), *Guide Application de la méthode MERMOS d'évaluation probabiliste de la fiabilité humaine pour les EPS de référence*, Note technique HT-54/02/020/A, EDF, France.

19. Desmares, E. & F. Cara (2000), *MERMOS: justifications théoriques*, HT-54/98/007/B, EDF, France.

20. Bieber, C., P. Le Bot, E. Desmares, F. Cara and J.-L. Bonnet (1998), MERMOS: EDF's new advanced HRA method, *Proceedings of the 4th International Conference on Probabilistic Safety Analysis and Management (PSAM4),* 13-18 September 1998, New York, United States.

21. Le Bot, P., C. Bieder and S. Vidal (2000), *Feedback from the actual implementation of the MERMOS method*, Proceedings of the 5th International Conference on Probabilistic Safety Analysis and Management (PSAM5), Osaka, Japan, 27 November-1 December 2000, Frontiers Science Series no. 34, Volume 3, Universal Academy Press, Tokyo, p. 1529-1534.

22. Kirwan, B., W.G. Gibson, J. Edmunds and R.J. Kennedy (2011), *Technical Basis for NARA, A Method for Human Error Quantification*, Corporate Risk Associates Ltd. (CRA), CRA-BEGL-POW-J032 Report No.1 Issue 7.

23. Edmunds, J., R.J. Kennedy, B. Kirwan & W.G. Gibson (2011), *A User Manual for the Nuclear Action Reliability Assessment (NARA) Human Error Quantification Technique*, CRA-BEGL-POW-J032 Report No. 2 Issue 8.

24. Gertman, D.I., H.S. Blackman, J.L. Marble, J.C. Byers and C.L. Smith (2005), *The SPAR-H Human Reliability Analysis Method*, NUREG/CR-6883, U.S. Nuclear Regulatory Commission, Washington, DC, United States.

25. Whaley, A.M., D.L. Kelly, R.L. Boring and W.J. Galyean (2011), *SPAR-H Step-by-Step Guidance*, INL/EXT-10-18533 Revision 2, Idaho National Laboratory, prepared for the U.S. Nuclear Regulatory Commission, United States.

26. Parry, G.W. et al (1992), *An Approach to the Analysis of Operator Actions in Probabilistic Risk Assessment*, EPRI TR-100259, Electric Power Research Institute, Palo Alto, United States.

27. Hollnagel, E. (1998), *Cognitive Reliability and Error Analysis Method (CREAM)*, Elsevier Science Ltd.

28. Spurgin, A.J. (1990) *Operator Reliability Experiments Using Power Plant Simulators,* Volume 2, EPRI NP-69373, Electric Power Research Institute, Palo Alto, United States.

29. Chu, T.L. et al (1994), *Evaluation of Potential Severe Accidents During Low Power and Shutdown Operations at Surrey, Unit 1: Analysis of Core Damage Frequency from Internal Events during Mid-Loop Operations*, Main Report (Chapters 7-12), NUREG/CR-6144, Volume 2, Part 1B, U.S. Nuclear Regulatory Commission, Washington, DC, United States.

30. Embrey, D.E., P.C. Humphreys, E.A. Rosa, B. Kirwan and K. Rea (1984), *SLIM-MAUD: An Approach to Assessing Human Error Probabilities Using Expert Judgement*, NUREG/CR-3518, U.S. Nuclear Regulatory Commission, Washington, DC, United States.

31. Kim, J. (2012), *The HuRECA-Based Human Reliability Analysis Procedure for Computer-based Control Room Actions,* 2012.2.

32. Kirwan, B., R. Kennedy, S. Taylor-Adams and B. Lambert (1997), *The validation of three human reliability quantification techniques – THERP, HEART and JHEDI: Part II – Results of validation exercise*, Applied Ergonomics, 28(1), 17-25.

33. Poucet, A. (1989*), The European Benchmark Exercise on Human Reliability Analysis*, Proceedings of the International Topical Meeting on Probability, Reliability and System Assessment (PSA'89): Pittsburgh, Pennsylvania, 2-7 April 1989, American Nuclear Society, Inc., La Grange, IL, pp. 103-110.

34. Williams, J.C. (1986), *HEART: A Proposed Method for Assessing and Reducing Human Error*, Proceedings of the 9th Advances in Reliability Technology Symposium, ARTS 86, University of Bradford, United Kingdom, 2-4 April 1986.

35. Reason, J. (1990), *Human Error*, Cambridge University Press.

36. American Society of Mechanical Engineers (2002), *Standard for Probabilistic Risk Assessment for Nuclear power Plant Applications*, ASME RA-S-2002, New York, NY, United States.

37. Hannaman, G.W. and A.J. Spurgin (1984), *Systematic Human Action Reliability Procedure (SHARP)*, EPRI NP 3583, Electric Power Research Institute, Palo Alto, United States.

APPENDIX 1. THE METHOD EVALUATION SCALE

Desirable attributes of HRA - Methods evaluation scale

Instructions to assessors

Indicate your evaluation of how well a method meets the requirements of an attribute by placing a tick in the relevant coloured box.

Provide a succinct justification for the rating you have allocated in the text box labelled justification.

You should complete an evaluation for each attribute and each sub-scale of an attribute.

If an attribute is not relevant to the method you are evaluating then record this in the justification box with an explanation as to why the attribute is not relevant.

Your evaluation should be based on the application of the method exactly as it is described in the method's user documentation. If there are modifications to the process which improve its application, but which are not formally recorded in an update to the method's documentation, then these can be noted in the justification box as potential improvements, but they should not be considered as part of the evaluation.

Within the method evaluation scale, a high rating (dark blue) indicates that the requirements of an attribute have been fully or largely met for the method's intended scope of application. An intermediate rating (medium blue) indicates that a method meets some, but not all of the requirements of the attribute. A low rating (light blue) indicates that the requirements of the attribute are not met or that no evidence is available in relation to the attribute for the method.

Construct validity	Essential	**Attribute 1** **Availability of information relating to the technical basis of the method** Information is provided on the technical basis of the method, in terms of its scientific underpinnings and data, in order to allow a judgement on the validity of the method to be made.		
		Comprehensive information on the technical basis and/or data underpinning the method is available and its application is discussed as part of the documentation of the method.	High	Justification
		The method provides references that allow the information forming the technical basis and/or the data underpinning it to be obtained.	Intermediate	
		The method does not provide sufficient information to allow its technical basis and underpinning data to be accessed for review.	Low	

Construct validity	Essential	**Attribute 2** **The Technical basis of the method (Theory)** The technical basis of the method is based upon, and does not contradict, a relevant body of scientific knowledge		
		The method operationalises a relevant model of human performance or system safety which has scientific acceptance.	High	Justification
		Elements of the method are inconsistent with an accepted scientific model of human performance or system safety.	Low	

Construct validity	Essential	**Attribute 3** **The technical basis of the method (Data)** Where the technical basis of the method is based on a dataset, the source of the data/information and its relevance for application in the nuclear industry should be demonstrated.		
		The data underlying the method are largely based on observations of actual or simulated task performance in nuclear industry tasks.	High	Justification
		The data underlying the method are based on expert judgement or observations of human performance for relevant tasks in a domain that is closely related to the nuclear industry e.g. other high hazard industries.	Intermediate	
		The data underlying the method are taken from tasks that are not related or relevant to nuclear industry tasks.	Low	

Construct validity	Highly desirable	**Attribute 4** **Internal consistency of the method** The method demonstrates internal consistency between the technical basis, the error definition, the PSFs and the qualitative and quantitative method steps		
		The qualitative and quantitative component parts of the method are theoretically compatible and form a coherent consistent whole.	High	Justification
		There are theoretical inconsistencies between the qualitative and/or quantitative components of the method.	Low	

Content Validity	Highly desirable	**Attribute 5** **Qualitative assessment** It is recognised good practice that HRA quantification is supported by qualitative analysis to develop an understanding of operator performance within the scenario that is being assessed. This attribute considers the extent to which the qualitative analysis stages of the HRA (e.g. task analysis and error identification) is directed or prescribed by the HRA method, beyond providing a set of performance shaping factors to be considered.		
		The method contains or prescribes a process for conducting qualitative assessment**.**	High	Justification
		The method includes a general statement indicating that a qualitative assessment should be provided, e.g. by referring to the use of task analysis**.**	Intermediate	
		The method does not make any reference to qualitative analysis**.**	Low	

Content validity	Essential	**Attribute 6** **Factors influencing human reliability considered by the method** The method should be quantitatively sensitive to a majority of accepted factors* (PSFs) that influence human reliability. *: There are pre-defined lists of PSFs/EPCs available throughout the literature and within HRA methods (typically 1st generation methods). This attribute does not seek to define such a list, so as to accommodate developments in human performance, system safety and accident analysis. The evaluation teams are asked to use professional judgment when considering this attribute. It is not expected that methods will be assigned a high rating (dark blue) if only a small number of factors are accommodated, but we do not prescribe an absolute number of factors that are required.		
		Sub-scale 1: Adequacy of PSFs.		
		The method requires qualitative assessment of a majority of accepted factors that affect human reliability (PSFs).	High	Justification
		The method does not consider a majority set of factors that affect human reliability.	Low	
		Sub-scale 2: Quantitative sensitivity.		
		The method is quantitatively sensitive to the effect of each individual PSF considered qualitatively.	High	Justification
		The method is not quantitatively sensitive to individual PSFs, but makes a single adjustment to the HEP based on the contribution of the overall combination of the PSFs considered.	Intermediate	
		The method is not quantitatively sensitive to PSFs.	Low	
		Sub-scale 3: Interaction between factors Typically HRA methods adopt a linear multiplicative combination of PSFs. It is recognised that some PSFs may interact in other ways, e.g. a step change in the effect of one PSF once a threshold has been reached on a second PSF, or where the effect of the combination of two PSFs is far greater than multiplicative relationship would predict or where one PSF has a triggering effect on other PSFs in a causal chain.		
		Interactions between PSFs are accounted for on the basis of knowledge of the relationship between specific PSFs.	High	Justification
		Combinations of PSF effects are accounted for using a simple linear model.	Intermediate	
		Interactions between or combination of PSF effects are not considered by the method.	Low	

Content validity	Essential	**Attribute 7** **Consideration of human error dependency** Modelling should include consideration of human error dependencies or common cause failures.		
		The method provides a procedure for identifying potential sources of dependence among Human Failure Events (HFEs) and/or sub-tasks of an HFE, and provides a method to derive conditional HEPs based on the systematic assessment of these sources of dependence.	High	Justification
		The method identifies potential sources of dependence, but does not provide a process for linking these sources of dependence to a quantified model for deriving conditional HEPs.	Intermediate	
		The method does not address dependencies and common cause mechanisms among tasks and sub-tasks.	Low	

Content validity	Essential	**Attribute 8** **Consideration of deviations and progressions in accident sequences** The method should provide a capability to accommodate: • Deviations from nominal accident scenarios due to: (A) Plant conditions: 1. Aleatory factors, such as sizes and locations of equipment failures and time sequences. 2. Complicating factors, such as coincident failures in control, instrumentation and support systems not normally modelled explicitly in PSA models. (B) Human failure scenarios due to crew organisational & operating practices that introduce opportunities to create new failure mechanisms. Fault progressions including consequential faults and accident sequences encompassing Level 1 and Level 2 PSA which may involve extended time sequences and degraded operating environments should also be accommodated.		
		Sub-scale 1 **Deviations**		
		The method provides for the qualitative and quantitative assessment of all the types of deviations in accident scenarios.	High	Justification
		The method provides for the qualitative assessment of human error during fault progressions, but does not provide for the derivation of HEPs in support of this assessment.	Intermediate	
		The method does not provide a means to deal with deviations in accident scenarios	Low	
		Sub-scale 2 **Fault progression**.		
		The method provides for the qualitative and quantitative assessment of human errors during fault progressions including level 1 to level 2 PSA fault progressions	High	Justification
		The method provides for the qualitative assessment of human error during fault progressions, but does not provide for the derivation of HEPs in support of this assessment.	Intermediate	
		The method does not provide for the qualitative and quantitative assessment of human errors during fault progressions.	Low	

Content validity	Highly desirable	**Attribute 9** **Consideration of cognitive error** The method should be sensitive to the factors that influence the diagnosis and decision making component of the response to an initiating event.		
		The method estimates the probability of cognitive error based on the assessment of a set of factors that are known to affect diagnosis and decision making performance	High	Justification
		The method uses a simple model such as a time reliability curve as the primary factor for estimating the probability of cognitive error**.**	Intermediate	
		The method provides no way of estimating the likelihood of cognitive error**.**	Low	

Content validity	Highly desirable	**Attribute 10** **Consideration of statistical uncertainty** The method should provide for statistical uncertainty analysis of derived human error probabilities.		
		The method derives uncertainty parameters from experience (either in-plant or from relevant simulator trials).	High	Justification
		The method provides generic uncertainty parameters, e.g. standardised error factors	Intermediate	
		The method provides no uncertainty parameters.	Low	

Content validity	Desirable	**Attribute 11** **Consideration of organisational issues** The method should consider the impact of organisational issues including safety-culture factors (attitudes and behaviours), and organisational process factors (e.g. command and control structures, conflicts of interest, communication and decision making protocols on human reliability).		
		Sub-scale 1 **Safety-culture factors** (attitudes and behaviours).		
		The method provides an adequate quantitative method to adjust HEPs based on an assessment of safety culture/safety climate**.**	High	Justification
		The method provides a qualitative means to assess safety culture/safety climate, but does not include a process to modify HEPs based on the assessment**.**	Intermediate	
		The method does not take into account safety culture factors**.**	Low	
		Sub-scale 2 **Process factors** (e.g. command and control structures, communication and decision making protocols on human reliability).		
		The method provides a quantitative method to assess process factors	High	Justification
		The method provides a qualitative means to assess process factors, but does not include a process to modify HEPs based on the assessment**.**	Intermediate	
		The method does not take into account process factors**.**	Low	

Empirical validity	Essential/Desirable	**Attribute 12** **Empirical validity** The method should demonstrate evidence of empirical validation exercises, peer review processes or community acceptance based on application and maturity.		
		Sub-scale1 **Statistical evidence**		
		The HEP estimates derived by the method have been shown to demonstrate good agreement with plant and /or simulator data for comparable tasks**.**	High	Justification
		The HEP estimates derived by the method have been shown to demonstrate good agreement with HEP estimates produced by other HRA methods for the same or comparable tasks**.**	Intermediate	
		The method has failed to derive comparable HEP estimates in tests of empirical validity or has not been subject to such assessments**.**	Low	
		Sub-scale 2 **Verification/Peer review**		
		The method has been subject to peer review by a team of recognised HRA experts, and the peer review comments have been incorporated to the development of the method**.**	High	Justification
		The method has been subject to peer review by a single, recognised HRA expert, and the comments have been incorporated to the development of the method**.**	Intermediate	
		The method has not been subject to independent peer review or the method has not been updated in response to peer review comments.	Low	
		Sub-scale 3 **Application/Maturity**		
		The method has been extensively applied, internationally, for five or more years**.**	High	Justification
		The method has been applied to a limited number of HRAs**.**	Intermediate	
		The method has not yet been applied to a HRA**.**	Low	

Reliability	Essential	**Attribute 13** **Computer models and software tools** If a method incorporates the use of a computer model or software tool to analyse a human action, A QA programme should be applied to ensure quality of the design and validity of the output.		
		A relevant, recognised/accepted international standard has been applied to the software design and verification of the computer based HRA method/tool**.**	High	Justification
		The design of the computer based HRA method/tool is based upon a documented QA process, which includes software verification**.**	Intermediate	
		There is no evidence that the design of the computer based HRA method/tool is based on a structured and validated software development or QA method that includes software verification**.**	Low	

Reliability	Highly desirable	**Attribute 14** **Reliability and traceability** The method should provide consistent qualitative and quantitative information for comparable scenarios within analysts and between analysts for similar scenarios. The method should also provide sufficient information to facilitate tracing estimates back to input assumptions.		
		Sub-scale 1 **Within analyst consistency/reliability**		
		A formal comparison, amenable to statistical analysis, has been undertaken to demonstrate that the same HRA analyst provides consistent answers for analyses made at different times for the same scenario**.**	High	Justification
		An informal comparison has been undertaken, which suggests good within analyst agreement for analyses made at different times**.**	Intermediate	
		There is no information available to suggest good within analyst agreement for analyses made at different times**.**	Low	
		Sub-scale 2 **Between analyst consistency/reliability**		
		A formal comparison, amenable to statistical analysis, has been undertaken to demonstrate that different HRA analysts provide consistent answers for the same scenario**.**	High	Justification
		An informal comparison has been undertaken, which suggests good between analyst agreement**.**	Intermediate	
		There is no information available to suggest good between analyst agreement**.**	Low	
		Sub-scale 3 **Traceability**		
		The method provides a procedure to ensure easy, complete traceability of the estimates of human performance in the HRA, such that an independent reviewer could trace back HEPs to relevant assumptions, models and data cited in the method**.**	High	Justification
		The HRA method itself does not provide a procedure for traceability, but there is sufficient information available about the method to facilitate traceability, and enable an independent reviewer to understand what was done**.**	Intermediate	
		There is insufficient information available to facilitate traceability**.**	Low	

Usability	Highly desirable	**Attribute 15** **Definition of method scope** The scope of the method should be clearly defined.		
		The scope of the method is clearly defined in a user manual and/or technical basis document**.**	High	Justification
		The scope of the method is described vaguely and some analyst judgement is required to determine its applicability to a particular human action/error**.**	Intermediate	
		The scope of the method is not defined**.**	Low	

Usability	Highly desirable	**Attribute 16** **Qualitative outputs** The method should produce qualitative outputs that are useful to inform human factors and safety management improvements at the plant		
		The method generates qualitative information to inform improvements to reduce the potential for human error that is explicitly related to each of the factors that are used in the method to derive an HEP.	High	Justification
		The method generates qualitative information to inform improvements to reduce the potential for human error, but this is not explicitly linked to each of the factors used in the derivation of HEPs.	Intermediate	
		The method does not generate qualitative information to inform improvements to reduce the potential for human error.	Low	

Usability	Highly desirable	**Attribute 17** **Qualitative uncertainty and quantitative conservatism** Methods should be able to reflect uncertainties related to qualitative information via conservatisms in the quantification process.		
		The method provides a mathematical procedure for adjusting the conservatism of the HEPs derived as a function of the level of certainty in the qualitative information collected during the assessment.	High	Justification
		The method provides a general caution on the need to adjust the conservatism of HEPs as a function of the level of certainty in the qualitative information collected, but does not provide a mathematical procedure for doing so.	Intermediate	
		The method does not address the issue of uncertainties in qualitative information and the impact of this on derived HEPs.	Low	

Usability	Desirable	**Attribute 18** **Availability of user documentation** The method should be supported by a detailed user documentation e.g., manual or instructions, which describes how the method should be applied.		
		The method contains user documentation that provides a detailed step-by-step procedure for all steps in the derivation of an HEP.	High	Justification
		The method contains user documentation that provides a high level description of how it is applied to derive HEPs, but not all elements of the method are detailed as step-by-step procedures.	Intermediate	
		The method provides only a high level description of its method of application and or data tables for the derivation of HEPs.	Low	

Usability	Desirable	**Attribute 19** **Use of limiting values** The method should provide limiting values. (Relevant Good Practice documents discuss limiting values that are used in member countries).		
		The method provides limiting values and advice on their application.	High	Justification
		The method provides advice on the need to limit claims on human performance but does not provide specific limiting values.	Intermediate	
		The method does not consider the use of limiting values.	Low	

Usability	Indifferent/Essential	**Attribute 20** **Resources** A comparative estimate of the resources (time, cost, utility demands, level of specialist training required, level and type of knowledge required) needed to apply the method in comparison with other HRA methods.		
		The estimated cost of and time required for applying the HRA method is less than or comparable with that of other HRA methods**.**	High	Justification
		The estimated cost of and time required for applying the HRA method is in excess of that required for application of other HRA methods**.**	Low	

APPENDIX 2 ATTRIBUTE EVALUATIONS FOR EACH METHOD

This appendix contains the attribute evaluation worksheets for each of the examined methods.

A2.1 Attribute Evaluations – THERP & ASEP

Desirable Attributes of HRA – Methods Evaluation Scale – THERP & ASEP

Instructions to assessors

Indicate your evaluation of how well a method meets the requirements of an attribute by placing a tick in the relevant coloured box.

Provide a succinct justification for the rating you have allocated in the text box labelled justification.

You should complete an evaluation for each attribute and each sub-scale of an attribute.

If an attribute is not relevant to the method you are evaluating then record this in the justification box with an explanation as to why the attribute is not relevant.

Your evaluation should be based on the application of the method exactly as it is described in the method's user documentation. If there are modifications to the process which improve its application, but which are not formally recorded in an update to the method's documentation, then these can be noted in the justification box as potential improvements, but they should not be considered as part of the evaluation.

Within the method evaluation scale, a high rating (dark blue) indicates that the requirements of an attribute have been fully or largely met for the method's intended scope of application. An intermediate rating (medium blue) indicates that a method meets some, but not all of the requirements of the attribute. A low rating (light blue) indicates that the requirements of the attribute are not met or that no evidence is available in relation to the attribute for the method.

Note: Where ASEP and THERP have different ratings,
- "T" is used to denote the THERP rating and
- "A" the ASEP rating.

Construct validity	Essential	**Attribute 1** **Availability of information relating to the technical basis of the method** Information is provided on the technical basis of the method, in terms of its scientific underpinnings and data, in order to allow a judgement on the validity of the method to be made.		THERP & ASEP
		Comprehensive information on the technical basis and/or data underpinning the method is available and its application is discussed as part of the documentation of the method.	X	Justification Extensive information on the THERP method is available in NUREG/CR-1278, Rev. 1* and its application process in NUREG/CR-2254**. Extensive information on the ASEP method is available in NUREG/CR-4772***.
		The method provides references that allow the information forming the technical basis and/or the data underpinning it to be obtained.		* Swain, A. D. and H.E. Guttmann, *Handbook of Human Reliability Analysis with Emphasis on Nuclear Power Plant Applications*, NUREG/CR-1278, Rev. 1. 1983, Sandia National Laboratories: Albuquerque, NM. ** Bell, B. J. and A.D. Swain, *A Procedure for Conducting a Human Reliability Analysis for Nuclear Power Plants*, NUREG/CR-2254. 1983, Sandia National Laboratories: Albuquerque, NM. (Available for download from *http://*prod.sandia.gov/techlib/access-control.cgi/1981/811655.pdf). *** Swain, A.D., *Accident Sequence Evaluation Program Human Reliability Analysis Procedure*, NUREG/CR-4772. 1987, Sandia National Laboratories: Albuquerque, NM
		The method does not provide sufficient information to allow its technical basis and underpinning data to be accessed for review.		

Construct validity	Essential	**Attribute 2** **The Technical basis of the method (Theory)** The technical basis of the method is based upon, and does not contradict, a relevant body of scientific knowledge		THERP & ASEP
		The method operationalises a relevant model of human performance or system safety which has scientific acceptance.	X	Justification Both methods are primarily based on a decompositional representation of human error probabilities, which has wide acceptance. The time/reliability correlation (T/RC) is similarly accepted, though neither is universally accepted.
		Elements of the method are inconsistent with an accepted scientific model of human performance or system safety.		

Construct validity	Essential	**Attribute 3** **The technical basis of the method (Data)** Where the technical basis of the method is based on a dataset, the source of the data/information and its relevance for application in the nuclear industry should be demonstrated.		THERP & ASEP
		The data underlying the method are largely based on observations of actual or simulated task performance in nuclear industry tasks.		Justification Many of the data in THERP and ASEP were judgements by Alan Swain based on observations made during nuclear weapons assembly and maintenance tasks.
		The data underlying the method are based on expert judgement or observations of human performance for relevant tasks in a domain that is closely related to the nuclear industry e.g. other high hazard industries.	X	
		The data underlying the method are taken from tasks that are not related or relevant to nuclear industry tasks.		

Construct validity	Highly desirable	**Attribute 4** **Internal consistency of the method** The method demonstrates internal consistency between the technical basis, the error definition, the PSFs and the qualitative and quantitative method steps		THERP & ASEP
		The qualitative and quantitative component parts of the method are theoretically compatible and form a coherent consistent whole.	X	Justification Both ASEP and THERP consist of two basic methods: the PSF-driven task analysis method and the time-based "cognitive" model. The PSF-driven model is entirely consistent with the qualitative components of the method. There is limited qualitative analysis for the T/RC.
		There are theoretical inconsistencies between the qualitative and/or quantitative components of the method.		

Content Validity	Highly desirable	**Attribute 5** **Qualitative assessment** It is recognised good practice that HRA quantification is supported by qualitative analysis to develop an understanding of operator performance within the scenario that is being assessed. This attribute considers the extent to which the qualitative analysis stages of the HRA (e.g. task analysis and error identification) is directed or prescribed by the HRA method, beyond providing a set of performance shaping factors to be considered.	THERP & ASEP	
		The method contains or prescribes a process for conducting qualitative assessment**.**	X	Justification In both methods, procedures are provided to describe the application of the methods. NUREG/CR-2254 and NUREG/CR-4772 provide guidance on how to perform the qualitative analysis and provide a specific process with instructions. However much of this process describes the evaluation process after the PRA systems analysts provide information.
		The method includes a general statement indicating that a qualitative assessment should be provided, e.g. by referring to the use of task analysis**.**		
		The method does not make any reference to qualitative analysis**.**		

Content validity	Essential	**Attribute 6** **Factors influencing human reliability considered by the method** The method should be quantitatively sensitive to a majority of accepted factors* (PSFs) that influence human reliability. *: There are pre-defined lists of PSFs/EPCs available throughout the literature and within HRA methods (typically 1st generation methods). This attribute does not seek to define such a list, so as to accommodate developments in human performance, system safety and accident analysis. The evaluation teams are asked to use professional judgment when considering this attribute. It is not expected that methods will be assigned a high rating (dark blue) if only a small number of factors are accommodated, but we do not prescribe an absolute number of factors that are required.	THERP & ASEP

Content validity	Essential	**Sub-scale 1** **Adequacy of PSFs**.		THERP & ASEP
		The method requires qualitative assessment of a majority of accepted factors that affect human reliability (PSFs).	T	Justification (Reviewers should identify the PSFs that are included in the method and judge the adequacy of this set for the intended application) The PSFs in THERP are the commonly used source of accepted PSFs in HRA, at least for most first generation methods. As well as the time available, the individual PSFs are too many to mention but generally cover the following areas: preparation, control of written materials and the structure of procedures; recollection of oral instructions; layout and types of displays; layout of controls on panels; layout of manual valves; use of tagging processes; effects of stress; level of checking; effects of walk-round checking. In ASEP, a limited number of PSFs are used for the nominal evaluation of post accident actions in addition to the use of time for the diagnostic steps. The most important are: the use of training, experience, or knowledge of the event, time between events, the type of action (step-by-step or dynamic) and the level of stress. Recovery factors are associated with post-action checking.
		The method does not consider a majority set of factors that affect human reliability.	A	
		Sub-scale 2 **Quantitative sensitivity**.		THERP & ASEP
		The method is quantitatively sensitive to the effect of each individual PSF considered qualitatively.	X	Justification The THERP and ASEP PSF tables used in the PSF-based method provide sensitivities for individual PSFs modelled.
		The method is not quantitatively sensitive to individual PSFs, but makes a single adjustment to the HEP based on the contribution of the overall combination of the PSFs considered.		
		The method is not quantitatively sensitive to PSFs.		

Content validity	Essential	**Sub-scale 3** **Interaction between factors** Typically HRA methods adopt a linear multiplicative combination of PSFs. It is recognised that some PSFs may interact in other ways, e.g. a step change in the effect of one PSF once a threshold has been reached on a second PSF, or where the effect of the combination of two PSFs is far greater than multiplicative relationship would predict or where one PSF has a triggering effect on other PSFs in a causal chain.		THERP & ASEP
		Interactions between PSFs are accounted for on the basis of knowledge of the relationship between specific PSFs.		Justification Interaction or combinations of PSFs are treated for the most part as independently linear.
		Combinations of PSF effects are accounted for using a simple linear model.	X	
		Interactions between or combination of PSF effects are not considered by the method.		

Content validity	Essential	**Attribute 7** **Consideration of human error dependency** Modelling should include consideration of human error dependencies or common cause failures.		THERP & ASEP
		The method provides a procedure for identifying potential sources of dependence among Human Failure Events (HFEs) and/or sub-tasks of an HFE, and provides a method to derive conditional HEPs based on the systematic assessment of these sources of dependence.	X	Justification There is extensive and detailed (very detailed) guidance for how to assess the dependencies (+ and -) between human actions in Chapter 10 of NUREG/CR-1278, Rev. 1.This does not apply to use of the TRC model. While the THERP method has five levels of dependence, ASEP models three.
		The method identifies potential sources of dependence, but does not provide a process for linking these sources of dependence to a quantified model for deriving conditional HEPs.		
		The method does not address dependencies and common cause mechanisms among tasks and sub-tasks.		

Content validity	Essential	**Attribute 8** **Consideration of deviations and progressions in accident sequences** The method should provide a capability to accommodate: • Deviations from nominal accident scenarios due to: (A) Plant conditions: 1. Aleatory factors, such as sizes and locations of equipment failures and time sequences. 2. Complicating factors, such as coincident failures in control, instrumentation and support systems not normally modelled explicitly in PSA models. (B) Human failure scenarios due to crew organisational & operating practices that introduce opportunities to create new failure mechanisms. • Fault progressions including consequential faults and accident sequences encompassing Level 1 and Level 2 PSA which may involve extended time sequences and degraded operating environments should also be accommodated.		THERP & ASEP
		Sub-scale 1 **Deviations**		THERP & ASEP
		The method provides for the qualitative and quantitative assessment of all the types of deviations in accident scenarios.		Justification Neither THERP nor ASEP provides direct way to consider deviations in scenarios. Rather the analyst should create these in the PRA's systems analysis tasks before applying the HRA method by developing new scenarios.
		The method provides for the qualitative assessment of human error during fault progressions, but does not provide for the derivation of HEPs in support of this assessment.		
		The method does not provide a means to deal with deviations in accident scenarios	X	

Content validity	Essential	**Sub-scale 2** **Fault progression**.		THERP & ASEP
		The method provides for the qualitative and quantitative assessment of human errors during fault progressions including level 1 to level 2 PSA fault progressions.		Justification THERP and ASEP provide no direct way to consider progressions in scenarios. Rather the analyst should create these before applying the method by developing new scenarios.
		The method provides for the qualitative assessment of human error during fault progressions, but does not provide for the derivation of HEPs in support of this assessment.		
		The method does not provide for the qualitative and quantitative assessment of human errors during fault progressions.	X	
Content validity	Highly desirable	**Attribute 9** **Consideration of cognitive error** The method should be sensitive to the factors that influence the diagnosis and decision making component of the response to an initiating event.		THERP & ASEP
		The method estimates the probability of cognitive error based on the assessment of a set of factors that are known to affect diagnosis and decision making performance		Justification THERP and ASEP both provide time/reliability (T/RC) models for diagnostic and decision-making tasks.
		The method uses a simple model such as a time reliability curve as the primary factor for estimating the probability of cognitive error**.**	X	
		The method provides no way of estimating the likelihood of cognitive error**.**		

Content validity	Highly desirable	**Attribute 10** **Consideration of statistical uncertainty** The method should provide for statistical uncertainty analysis of derived human error probabilities.		THERP & ASEP
		The method derives uncertainty parameters from experience (either in-plant or from relevant simulator trials).		Justification THERP and ASEP present median and error factors for all HEPs. These are based on the judgment of the authors of the methods.
		The method provides generic uncertainty parameters, e.g. standardised error factors	X	
		The method provides no uncertainty parameters.		

Content validity	Desirable	**Attribute 11** **Consideration of organisational issues** The method should consider the impact of organisational issues including safety-culture factors (attitudes and behaviours), and organisational process factors (e.g. command and control structures, conflicts of interest, communication and decision making protocols on human reliability).		THERP & ASEP
		Sub-scale 1 **Safety-culture factors** (attitudes and behaviours).		THERP & ASEP
		The method provides an adequate quantitative method to adjust HEPs based on an assessment of safety culture/safety climate**.**		Justification Table 3-2 in NUREG/CR-1278 lists numerous factors that may be considered cultural but provides no means for incorporating them in the assessment. There is no consideration reported for ASEP.
		The method provides a qualitative means to assess safety culture/safety climate, but does not include a process to modify HEPs based on the assessment**.**		
		The method does not take into account safety culture factors**.**	X	

Content validity	Desirable	**Sub-scale 2** **Process factors** (e.g. command and control structures, communication and decision making protocols on human reliability).		THERP & ASEP
		The method provides a quantitative method to assess process factors		Justification The documentation of THERP discusses process factors and some are represented in the evaluation of administration controls. [This ranking reflects the limited set of controls modelled.] There is no consideration reported for ASEP.
		The method provides a qualitative means to assess process factors, but does not include a process to modify HEPs based on the assessment**.**	T	
		The method does not take into account process factors**.**	A	

Empirical validity	Essential/Desirable	**Attribute 12** **Empirical validity** The method should demonstrate evidence of empirical validation exercises, peer review processes or community acceptance based on application and maturity.		THERP & ASEP
		Sub-scale1 **Statistical evidence**		THERP & ASEP
		The HEP estimates derived by the method have been shown to demonstrate good agreement with plant and /or simulator data for comparable tasks**.**	T	Justification The evaluation by Kirwan et al,. (*Applied Ergonomics* Vol. 28. No. I, pp. 17-25. 1997) indicates a good agreement with a variety of task data for THERP, but most were not from NPP applications. No such evaluations have been explicitly made for ASEP.
		The HEP estimates derived by the method have been shown to demonstrate good agreement with HEP estimates produced by other HRA methods for the same or comparable tasks**.**		
		The method has failed to derive comparable HEP estimates in tests of empirical validity or has not been subject to such assessments**.**	A	

Empirical validity	Essential/Desirable	**Sub-scale 2** **Verification/Peer review**		THERP & ASEP
		The method has been subject to peer review by a team of recognised HRA experts, and the peer review comments have been incorporated to the development of the method**.**	T	Justification There have been numerous reviews of THERP. The Rev. 1 document was modified to take account of comments made on the original THERP documentation. ASEP was reviewed as part of the NRC's evaluation of HRA methods vs. their good practices guide in NUREG-1842. No changes were made to the method. The only other review of ASEP *per se* is an unpublished review by Wreathall (2009) for NRC, and is therefore considered not reviewed.
		The method has been subject to peer review by a single, recognised HRA expert, and the comments have been incorporated to the development of the method**.**		
		The method has not been subject to independent peer review or the method has not been updated in response to peer review comments.	A	
		Sub-scale 3 **Application/Maturity**		THERP & ASEP
		The method has been extensively applied, internationally, for five or more years**.**	X	Justification THERP must be the most often used (or claimed to be used) HRA method. (However it has been noted that many applications do not follow completely the method as documented but typically make shortcuts for reasons of "efficiency".) ASEP has been used (or cited as being used) in several industry-performed PRAs in the USA and elsewhere.
		The method has been applied to a limited number of HRAs**.**		
		The method has not yet been applied to a HRA**.**		

Reliability	Essential	**Attribute 13** **Computer models and software tools** If a method incorporates the use of a computer model or software tool to analyse a human action, A QA programme should be applied to ensure quality of the design and validity of the output.		THERP & ASEP
		A relevant, recognised/accepted international standard has been applied to the software design and verification of the computer based HRA method/tool**.**		Justification N/A. No computer based models are used in the typical application of THERP. (It is noted that a version of THERP is built into the EPRI Calculator)
		The design of the computer based HRA method/tool is based upon a documented QA process, which includes software verification**.**	A	The propagation of ASEP uncertainty analysis is available as a computer program whose code is presented in NUREG/CR-4772. The extent of its use is not known.
		There is no evidence that the design of the computer based HRA method/tool is based on a structured and validated software development or QA method that includes software verification**.**		

Reliability	Highly desirable	**Attribute 14** **Reliability and traceability** The method should provide consistent qualitative and quantitative information for comparable scenarios within analysts and between analysts for similar scenarios. The method should also provide sufficient information to facilitate tracing estimates back to input assumptions.		THERP & ASEP
		Sub-scale 1. **Within analyst consistency/reliability**		THERP & ASEP
		A formal comparison, amenable to statistical analysis, has been undertaken to demonstrate that the same HRA analyst provides consistent answers for analyses made at different times for the same scenario**.**		Justification None known.
		An informal comparison has been undertaken, which suggests good within analyst agreement for analyses made at different times**.**		
		There is no information available to suggest good within analyst agreement for analyses made at different times**.**	X	

Reliability	Highly desirable	**Sub-scale 2** **Between analyst consistency/reliability**		THERP & ASEP
		A formal comparison, amenable to statistical analysis, has been undertaken to demonstrate that different HRA analysts provide consistent answers for the same scenario**.**		Justification Studies performed have shown that different analysts may get similar quantitative answers but have very different qualitative analyses (e.g. Kirwan et al 1997). The U.S. HRA Empirical Study [Ref. 9] addressed the consistency/reliability of analyses performed by different analysis teams with the ASEP method (as well as other analysis teams applying CBDT+HCR/ORE, SPAR-H, and ATHEANA). Although the results with ASEP (as well as with the other methods) found some consistency in the HEPs obtained for the HFEs examined in the study, a detailed comparison of the HRAs of the HFEs found significant differences in the qualitative findings used by the analysis teams to estimate the HEPs. Furthermore, there were also significant differences in the assessed contribution of the diagnosis/decision and execution components of the HEPs. Consequently, in spite of the limitations of the U.S. HRA Empirical Study, its results suggest that the between-analyst consistency is superficial. In conclusion, this information does not suggest good between-analyst agreement.
		An informal comparison has been undertaken, which suggests good between analyst agreement**.**		
		There is no information available to suggest good between analyst agreement**.**	X	
		Sub-scale 3 **Traceability**		THERP & ASEP
		The method provides a procedure to ensure easy, complete traceability of the estimates of human performance in the HRA, such that an independent reviewer could trace back HEPs to relevant assumptions, models and data cited in the method**.**	X	Justification When applied as documented, ASEP and THERP results can be easily traced and assessed by reviewers
		The HRA method itself does not provide a procedure for traceability, but there is sufficient information available about the method to facilitate traceability, and enable an independent reviewer to understand what was done**.**		
		There is insufficient information available to facilitate traceability**.**		

Usability	Highly desirable	**Attribute 15** **Definition of method scope** The scope of the method should be clearly defined.		THERP & ASEP
		The scope of the method is clearly defined in a user manual and/or technical basis document**.**	X	Justification Both methods are aimed at Level 1 PRAs actions by humans, both pre-accident and post-accident.
		The scope of the method is described vaguely and some analyst judgement is required to determine its applicability to a particular human action/error**.**		
		The scope of the method is not defined**.**		

Usability	Highly desirable	**Attribute 16** **Qualitative outputs** The method should produce qualitative outputs that are useful to inform human factors and safety management improvements at the plant.		THERP & ASEP
		The method generates qualitative information to inform improvements to reduce the potential for human error that is explicitly related to each of the factors that are used in the method to derive an HEP.	X	Justification By the numerical evaluation of the PSFs, it is possible to identify those areas of human performance can be improved and what would be the effects. The use of importance measures in the PRA quantification allows the analyst to identify what areas need improvement first. However many of the PSFs modelled in THERP and (especially) ASEP may be no longer critical risk issues in modern nuclear power plants.
		The method generates qualitative information to inform improvements to reduce the potential for human error, but this is not explicitly linked to each of the factors used in the derivation of HEPs.		
		The method does not generate qualitative information to inform improvements to reduce the potential for human error.		

Usability	Highly desirable	**Attribute 17** **Qualitative uncertainty and quantitative conservatism** Methods should be able to reflect uncertainties related to qualitative information via conservatisms in the quantification process.		THERP & ASEP
		The method provides a mathematical procedure for adjusting the conservatism of the HEPs derived as a function of the level of certainty in the qualitative information collected during the assessment.		Justification The methods do not discuss the effects of uncertainties in input information.
		The method provides a general caution on the need to adjust the conservatism of HEPs as a function of the level of certainty in the qualitative information collected, but does not provide a mathematical procedure for doing so.		
		The method does not address the issue of uncertainties in qualitative information and the impact of this on derived HEPs.	X	

Usability	Desirable	**Attribute 18** **Availability of user documentation** The method should be supported by a detailed user documentation e.g., manual or instructions, which describes how the method should be applied.		THERP & ASEP
		The method contains user documentation that provides a detailed step-by-step procedure for all steps in the derivation of an HEP.	X	Justification Both the THERP manual and the user guide (listed earlier) provide more than sufficient details on use of the method. This is similarly true for ASEP.
		The method contains user documentation that provides a high level description of how it is applied to derive HEPs, but not all elements of the method are detailed as step-by-step procedures.		
		The method provides only a high level description of its method of application and or data tables for the derivation of HEPs.		

Usability	Desirable	**Attribute 19** **Use of limiting values** The method should provide limiting values. (Relevant Good Practice documents discuss limiting values that are used in member countries).		THERP & ASEP
		The method provides limiting values and advice on their application.		Justification While the use of the dependence models in THERP and ASEP for the PSF analyses effectively limits the lower bound values, there is no explicit bound value provided. In the case of the T/RC there is an effective lower bound cut-off.
		The method provides advice on the need to limit claims on human performance but does not provide specific limiting values.	X	
		The method does not consider the use of limiting values.		

Usability	Indifferent/Essential	**Attribute 20** **Resources** A comparative estimate of the resources (time, cost, utility demands, level of specialist training required, level and type of knowledge required) needed to apply the method in comparison with other HRA methods.		THERP & ASEP
		The estimated cost of and time required for applying the HRA method is less than or comparable with that of other HRA methods**.**	A	Justification THERP, <u>when applied as documented</u>, would take more time than other methods, including the usual simplified version that most users apply. In the case of ASEP, the method takes less time than most other methods. Both methods require reasonable but not excessive time and resources to be provided by utilities, including access to control panels, procedures and discussions with operators and trainers. ASEP requires less than THERP. The method as laid out in the documentation can be easily followed, but the interface with the PSA models requires some knowledge of nuclear plant technology and safety issues. This reviewer is not aware of training courses in THERP or ASEP these days. Self-training and apprenticeships with experienced analysts are more common.
		The estimated cost of and time required for applying the HRA method is in excess of that required for application of other HRA methods**.**	T	

A2.2 Attribute Evaluations – Enhanced Bayesian THERP

Desirable Attributes of HRA – Methods Evaluation Scale – Enhanced Bayesian THERP

Instructions to assessors

Indicate your evaluation of how well a method meets the requirements of an attribute by placing a tick in the relevant coloured box.

Provide a succinct justification for the rating you have allocated in the text box labelled justification.

You should complete an evaluation for each attribute and each sub-scale of an attribute.

If an attribute is not relevant to the method you are evaluating then record this in the justification box with an explanation as to why the attribute is not relevant.

Your evaluation should be based on the application of the method exactly as it is described in the method's user documentation. If there are modifications to the process which improve its application, but which are not formally recorded in an update to the method's documentation, then these can be noted in the justification box as potential improvements, but they should not be considered as part of the evaluation.

Within the method evaluation scale, a high rating (dark blue) indicates that the requirements of an attribute have been fully or largely met for the method's intended scope of application. An intermediate rating (medium blue) indicates that a method meets some, but not all of the requirements of the attribute. A low rating (light blue) indicates that the requirements of the attribute are not met or that no evidence is available in relation to the attribute for the method.

Construct validity	Essential	**Attribute 1** **Availability of information relating to the technical basis of the method** Information is provided on the technical basis of the method, in terms of its scientific underpinnings and data, in order to allow a judgement on the validity of the method to be made.		Enhanced Bayesian THERP
		Comprehensive information on the technical basis and/or data underpinning the method is available and its application is discussed as part of the documentation of the method.		Justification There is little in the way of formal documentation of the method to allow judgement of the technical basis, though confidential descriptions are available in PRAs that have used the method. Conference papers and research reports providing an overview of the method are publicly available.
		The method provides references that allow the information forming the technical basis and/or the data underpinning it to be obtained.	X	
		The method does not provide sufficient information to allow its technical basis and underpinning data to be accessed for review.		

Construct validity	Essential	**Attribute 2** **The Technical basis of the method (Theory)** The technical basis of the method is based upon, and does not contradict, a relevant body of scientific knowledge**.**		Enhanced Bayesian THERP
		The method operationalises a relevant model of human performance or system safety which has scientific acceptance.	X	Justification The method broadly is consistent with the PSF type of HRA method. This is inferred from the PSFs used and their relationship with the underlying THERP T/RC.
		Elements of the method are inconsistent with an accepted scientific model of human performance or system safety.		

Construct validity	Essential	**Attribute 3** **The technical basis of the method (Data)** Where the technical basis of the method is based on a dataset, the source of the data/information and its relevance for application in the nuclear industry should be demonstrated.		Enhanced Bayesian THERP
		The data underlying the method are largely based on observations of actual or simulated task performance in nuclear industry tasks.		Justification The basic data for this method are derived from the THERP T/RC; however, there are unexplained deviations from the basic THERP T/RC . The effectiveness of the PSFs is largely judgemental on the part of the analysts, though guidance is provided from the early applications as exemplars for future analyses.
		The data underlying the method are based on expert judgement or observations of human performance for relevant tasks in a domain that is closely related to the nuclear industry e.g. other high hazard industries.	X	
		The data underlying the method are taken from tasks that are not related or relevant to nuclear industry tasks.		

Construct validity	Highly desirable	**Attribute 4** **Internal consistency of the method** The method demonstrates internal consistency between the technical basis, the error definition, the PSFs and the qualitative and quantitative method steps		Enhanced Bayesian THERP
		The qualitative and quantitative component parts of the method are theoretically compatible and form a coherent consistent whole.	X	Justification The combined use of the THERP T/RC and the PSFs as adjustments to its point estimates is a coherent approach, both qualitatively and quantitatively.
		There are theoretical inconsistencies between the qualitative and/or quantitative components of the method.		

Content Validity	Highly desirable	**Attribute 5** **Qualitative assessment** It is recognised good practice that HRA quantification is supported by qualitative analysis to develop an understanding of operator performance within the scenario that is being assessed. This attribute considers the extent to which the qualitative analysis stages of the HRA (e.g. task analysis and error identification) is directed or prescribed by the HRA method, beyond providing a set of performance shaping factors to be considered.		Enhanced Bayesian THERP
		The method contains or prescribes a process for conducting qualitative assessment**.**		Justification The documentation generally refers to the use of typical HRA modelling methods.
		The method includes a general statement indicating that a qualitative assessment should be provided, e.g. by referring to the use of task analysis**.**	X	
		The method does not make any reference to qualitative analysis**.**		

Content validity	Essential	**Attribute 6** **Factors influencing human reliability considered by the method** The method should be quantitatively sensitive to a majority of accepted factors* (PSFs) that influence human reliability. *: There are pre-defined lists of PSFs/EPCs available throughout the literature and within HRA methods (typically 1st generation methods). This attribute does not seek to define such a list, so as to accommodate developments in human performance, system safety and accident analysis. The evaluation teams are asked to use professional judgment when considering this attribute. It is not expected that methods will be assigned a high rating (dark blue) if only a small number of factors are accommodated, but we do not prescribe an absolute number of factors that are required.		Enhanced Bayesian THERP
		Sub-scale 1 **Adequacy of PSFs**		Enhanced Bayesian THERP
		The method requires qualitative assessment of a majority of accepted factors that affect human reliability (PSFs).	X	Justification (Reviewers should identify the PSFs that are included in the method and judge the adequacy of this set for the intended application) The combination of PSFs and the TRC seem to cover most post-initiator human factors concerns. The PSFs used are: K1: Quality and relevance of procedures. K2: Quality and relevance of training. K3: Quality and relevance of feedback from process (MMI). K4: Mental load (stress) in the situation. K5: Need for coordination and communication.
		The method does not consider a majority set of factors that affect human reliability.		
		Sub-scale 2 **Quantitative sensitivity**		Enhanced Bayesian THERP
		The method is quantitatively sensitive to the effect of each individual PSF considered qualitatively.	X	Justification The effects of each PSF are analysed individually for their effect on the T/RC.
		The method is not quantitatively sensitive to individual PSFs, but makes a single adjustment to the HEP based on the contribution of the overall combination of the PSFs considered.		
		The method is not quantitatively sensitive to PSFs.		

Content validity	Essential	**Sub-scale 3** **Interaction between factors** Typically HRA methods adopt a linear multiplicative combination of PSFs. It is recognised that some PSFs may interact in other ways, e.g. a step change in the effect of one PSF once a threshold has been reached on a second PSF, or where the effect of the combination of two PSFs is far greater than multiplicative relationship would predict or where one PSF has a triggering effect on other PSFs in a causal chain.		Enhanced Bayesian THERP
		Interactions between PSFs are accounted for on the basis of knowledge of the relationship between specific PSFs.		Justification Combinations of the effects of PSFs are calculated using Bayesian mathematical equations, but the effects of individual PSFs are considered separately.
		Combinations of PSF effects are accounted for using a simple linear model.	X	
		Interactions between or combination of PSF effects are not considered by the method.		

Content validity	Essential	**Attribute 7** **Consideration of human error dependency** Modelling should include consideration of human error dependencies or common cause failures.		Enhanced Bayesian THERP
		The method provides a procedure for identifying potential sources of dependence among Human Failure Events (HFEs) and/or sub-tasks of an HFE, and provides a method to derive conditional HEPs based on the systematic assessment of these sources of dependence.	X	Justification Dependencies are identified both in the qualitative analysis phase and in the cut-set investigation phase. Full dependence is suggested for multiple operator actions in the same minimal cut set.
		The method identifies potential sources of dependence, but does not provide a process for linking these sources of dependence to a quantified model for deriving conditional HEPs.		
		The method does not address dependencies and common cause mechanisms among tasks and sub-tasks.		

Content validity | Essential

Attribute 8
Consideration of deviations and progressions in accident sequences

Enhanced Bayesian THERP

The method should provide a capability to accommodate:

- Deviations from nominal accident scenarios due to:
  - (A) Plant conditions:
    1. Aleatory factors, such as sizes and locations of equipment failures and time sequences.
    2. Complicating factors, such as coincident failures in control, instrumentation and support systems not normally modelled explicitly in PSA models.
  - (B) Human failure scenarios due to crew organisational & operating practices that introduce opportunities to create new failure mechanisms.
- Fault progressions including consequential faults and accident sequences encompassing Level 1 and Level 2 PSA which may involve extended time sequences and degraded operating environments should also be accommodated.

Sub-scale 1. Deviations

The method provides for the qualitative and quantitative assessment of all the types of deviations in accident scenarios.		Justification The method does not provide any explicit means to identify deviations in accident sequences.
The method provides for the qualitative assessment of human error during fault progressions, but does not provide for the derivation of HEPs in support of this assessment.		
The method does not provide a means to deal with deviations in accident scenarios	X	

Sub-scale 2
Fault progression

Enhanced Bayesian THERP

The method provides for the qualitative and quantitative assessment of human errors during fault progressions including level 1 to level 2 PSA fault progressions.		Justification The underlying T/RC is based on the time operators have to respond to prevent core damage from occurring. In principle the same kinds of PSFs could be used for level 2 analyses. Conceptually the method could be used into level 2 events but there is no support for it at present.
The method provides for the qualitative assessment of human error during fault progressions, but does not provide for the derivation of HEPs in support of this assessment.		

		The method does not provide for the qualitative and quantitative assessment of human errors during fault progressions.	X	

Content validity	Highly desirable	**Attribute 9** **Consideration of cognitive error** The method should be sensitive to the factors that influence the diagnosis and decision making component of the response to an initiating event.		Enhanced Bayesian THERP
		The method estimates the probability of cognitive error based on the assessment of a set of factors that are known to affect diagnosis and decision making performance	X	Justification The PSFs used are considered appropriate for the estimation of failures in cognition. The method is therefore more appropriate than just the use of the T/RC.
		The method uses a simple model such as a time reliability curve as the primary factor for estimating the probability of cognitive error**.**		
		The method provides no way of estimating the likelihood of cognitive error**.**		

Content validity	Highly desirable	**Attribute 10** **Consideration of statistical uncertainty** The method should provide for statistical uncertainty analysis of derived human error probabilities.		Enhanced Bayesian THERP
		The method derives uncertainty parameters from experience (either in-plant or from relevant simulator trials).		Justification The method explicitly allows for the assessment of uncertainties but these are based on judgment rather than actual data.
		The method provides generic uncertainty parameters, e.g. standardised error factors	X	
		The method provides no uncertainty parameters.		

Content validity	Desirable	**Attribute 11** **Consideration of organisational issues** The method should consider the impact of organisational issues including safety-culture factors (attitudes and behaviours), and organisational process factors (e.g. command and control structures, conflicts of interest, communication and decision making protocols on human reliability).		Enhanced Bayesian THERP
		Sub-scale 1 **Safety-culture factors** (attitudes and behaviours)		Enhanced Bayesian THERP
		The method provides an adequate quantitative method to adjust HEPs based on an assessment of safety culture/safety climate**.**		Justification None of the PSFs used nor the T/RC represent any safety culture factors.
		The method provides a qualitative means to assess safety culture/safety climate, but does not include a process to modify HEPs based on the assessment**.**		
		The method does not take into account safety culture factors**.**	X	
		Sub-scale 2 **Process factors** (e.g. command and control structures, communication and decision making protocols on human reliability).		Enhanced Bayesian THERP
		The method provides a quantitative method to assess process factors		Justification None of the PSFs used nor the T/RC represent any process factors, though one PSF requires consideration of the need for co-ordination and communication. There appears to be no assessment of their availability or quality.
		The method provides a qualitative means to assess process factors, but does not include a process to modify HEPs based on the assessment**.**		
		The method does not take into account process factors**.**	X	

Empirical validity	Essential/Desirable	**Attribute 12** **Empirical validity** The method should demonstrate evidence of empirical validation exercises, peer review processes or community acceptance based on application and maturity.		Enhanced Bayesian THERP
		Sub-scale1 **Statistical evidence**		Enhanced Bayesian THERP
		The HEP estimates derived by the method have been shown to demonstrate good agreement with plant and /or simulator data for comparable tasks**.**		Justification It is understood that there are close agreements with data gathered in the International Benchmarking HRA study documented in NUREG/IA-0216 Volumes 1-3. However due to the non-statistical treatment of the data generated by the international empirical study, it is not considered to provide evidence in relation to this attribute in this study.
		The HEP estimates derived by the method have been shown to demonstrate good agreement with HEP estimates produced by other HRA methods for the same or comparable tasks**.**		
		The method has failed to derive comparable HEP estimates in tests of empirical validity or has not been subject to such assessments**.**	X	
		Sub-scale 2 **Verification/Peer review**		Enhanced Bayesian THERP
		The method has been subject to peer review by a team of recognised HRA experts, and the peer review comments have been incorporated to the development of the method**.**	X	Justification There have been several reviews of the method by regulatory bodies in Scandinavia which is the basis for the assignment of the high rating. The method is also part of the International Benchmarking Study and the Nordic/German HRA method comparison.
		The method has been subject to peer review by a single, recognised HRA expert, and the comments have been incorporated to the development of the method**.**		
		The method has not been subject to independent peer review or the method has not been updated in response to peer review comments.		

Empirical validity	Essential/Desirable	**Sub-scale 3** **Application/Maturity**		Enhanced Bayesian THERP
		The method has been extensively applied, internationally, for five or more years**.**		Justification The method has been applied in three PRAs and the International Benchmarking Study.
		The method has been applied to a limited number of HRAs**.**	X	
		The method has not yet been applied to a HRA**.**		

Reliability	Essential	**Attribute 13** **Computer models and software tools** If a method incorporates the use of a computer model or software tool to analyse a human action, A QA programme should be applied to ensure quality of the design and validity of the output.		Enhanced Bayesian THERP
		A relevant, recognised/accepted international standard has been applied to the software design and verification of the computer based HRA method/tool**.**		Justification N/A. The method uses off-the-shelf software (MS Excel).
		The design of the computer based HRA method/tool is based upon a documented QA process, which includes software verification**.**		
		There is no evidence that the design of the computer based HRA method/tool is based on a structured and validated software development or QA method that includes software verification**.**		

Reliability	Highly desirable	**Attribute 14** **Reliability and traceability** The method should provide consistent qualitative and quantitative information for comparable scenarios within analysts and between analysts for similar scenarios. The method should also provide sufficient information to facilitate tracing estimates back to input assumptions.		Enhanced Bayesian THERP
		Sub-scale 1 **Within analyst consistency/reliability**		Enhanced Bayesian THERP
		A formal comparison, amenable to statistical analysis, has been undertaken to demonstrate that the same HRA analyst provides consistent answers for analyses made at different times for the same scenario**.**		Justification No such evaluation has been mentioned in the available documentation.
		An informal comparison has been undertaken, which suggests good within analyst agreement for analyses made at different times**.**		
		There is no information available to suggest good within analyst agreement for analyses made at different times**.**	X	
		Sub-scale 2 **Between analyst consistency/reliability**		Enhanced Bayesian THERP
		A formal comparison, amenable to statistical analysis, has been undertaken to demonstrate that different HRA analysts provide consistent answers for the same scenario**.**		Justification It is noted that a team should undertake the assessment of PSFs and the Bayesian process combines their assessments. Hence the method can accommodate between-analyst differences. However there has been no formal test of between analyst reliability.
		An informal comparison has been undertaken, which suggests good between analyst agreement**.**		
		There is no information available to suggest good between analyst agreement**.**	X	

Reliability	Highly desirable	**Sub-scale 3** **Traceability**		Enhanced Bayesian THERP
		The method provides a procedure to ensure easy, complete traceability of the estimates of human performance in the HRA, such that an independent reviewer could trace back HEPs to relevant assumptions, models and data cited in the method**.**		Justification Whilst, the method makes clear each step in the analysis through the use of identified scales for the PSF ratings, the use of the T/RC and the results not all steps are provided in sufficient detail.
		The HRA method itself does not provide a procedure for traceability, but there is sufficient information available about the method to facilitate traceability, and enable an independent reviewer to understand what was done**.**	X	
		There is insufficient information available to facilitate traceability**.**		

Usability	Highly desirable	**Attribute 15** **Definition of method scope** The scope of the method should be clearly defined.		Enhanced Bayesian THERP
		The scope of the method is clearly defined in a user manual and/or technical basis document**.**	X	Justification The scope of the method is clearly defined in the available literature and is aimed at post-initiating event human actions.
		The scope of the method is described vaguely and some analyst judgement is required to determine its applicability to a particular human action/error**.**		
		The scope of the method is not defined**.**		

Usability	Highly desirable	**Attribute 16** **Qualitative outputs** The method should produce qualitative outputs that are useful to inform human factors and safety management improvements at the plant.		Enhanced Bayesian THERP
		The method generates qualitative information to inform improvements to reduce the potential for human error that is explicitly related to each of the factors that are used in the method to derive an HEP.		Justification The ratings of each PSF identify effectively what areas of human performance (within the scope of the model) need to be improved, and the rating scale suggests what kinds of changes need to be made. However, no specific corrections are suggested.
		The method generates qualitative information to inform improvements to reduce the potential for human error, but this is not explicitly linked to each of the factors used in the derivation of HEPs.	X	
		The method does not generate qualitative information to inform improvements to reduce the potential for human error.		

Usability	Highly desirable	**Attribute 17** **Qualitative uncertainty and quantitative conservatism** Methods should be able to reflect uncertainties related to qualitative information via conservatisms in the quantification process.		Enhanced Bayesian THERP
		The method provides a mathematical procedure for adjusting the conservatism of the HEPs derived as a function of the level of certainty in the qualitative information collected during the assessment.		Justification The method provides limited guidance on how to accommodate uncertainties associated with input information.
		The method provides a general caution on the need to adjust the conservatism of HEPs as a function of the level of certainty in the qualitative information collected, but does not provide a mathematical procedure for doing so.	X	
		The method does not address the issue of uncertainties in qualitative information and the impact of this on derived HEPs.		

Usability	Desirable	**Attribute 18** **Availability of user documentation** The method should be supported by a detailed user documentation e.g., manual or instructions, which describes how the method should be applied.		Enhanced Bayesian THERP
		The method contains user documentation that provides a detailed step-by-step procedure for all steps in the derivation of an HEP.		Justification The method is described in a series of case studies in papers and reports. These are generally sufficient to understand the process of the method but are not explicitly a user manual.
		The method contains user documentation that provides a high level description of how it is applied to derive HEPs, but not all elements of the method are detailed as step-by-step procedures.	X	
		The method provides only a high level description of its method of application and or data tables for the derivation of HEPs.		

Usability	Desirable	**Attribute 19** **Use of limiting values** The method should provide limiting values. (Relevant Good Practice documents discuss limiting values that are used in member countries).		Enhanced Bayesian THERP
		The method provides limiting values and advice on their application.		Justification The use of the T/RC limits the values of HEPs that can be predicted. However, the use of multiple PSFs that are rated very good, could lead to very low probabilities. There appears to be no prohibition or advice concerning this situation.
		The method provides advice on the need to limit claims on human performance but does not provide specific limiting values.		
		The method does not consider the use of limiting values.	X	

Usability	Indifferent/Essential	**Attribute 20** **Resources** A comparative estimate of the resources (time, cost, utility demands, level of specialist training required, level and type of knowledge required) needed to apply the method in comparison with other HRA methods.		Enhanced Bayesian THERP
		The estimated cost of and time required for applying the HRA method is less than or comparable with that of other HRA methods**.**	X	Justification This method is expected to require resources and effort typically the same as other PSF-based methods.
		The estimated cost of and time required for applying the HRA method is in excess of that required for application of other HRA methods**.**		This method is not likely to require major demands on utility resources, though it must be recognised that utility personnel (operators and trainers) should be part of any HRA study, to provide operating experience that is missed by analysts without such experience. The evaluation of the PSFs should be within the skill set of experienced HRA analysts, though training in the specific anchor points for the PSF ratings is suggested. In most cases to date the method has been applied by its developers. Training to utility staff is normally provided during the application process. However, it is judged that any training to external users would not be onerous.

A2.3 Attribute Evaluations – ATHEANA

Desirable Attributes of HRA – Methods Evaluation Scale – ATHEANA

Instructions to assessors

Indicate your evaluation of how well a method meets the requirements of an attribute by placing a tick in the relevant coloured box.

Provide a succinct justification for the rating you have allocated in the text box labelled justification.

You should complete an evaluation for each attribute and each sub-scale of an attribute.

If an attribute is not relevant to the method you are evaluating then record this in the justification box with an explanation as to why the attribute is not relevant.

Your evaluation should be based on the application of the method exactly as it is described in the method's user documentation. If there are modifications to the process which improve its application, but which are not formally recorded in an update to the method's documentation, then these can be noted in the justification box as potential improvements, but they should not be considered as part of the evaluation.

Within the method evaluation scale, a high rating (dark blue) indicates that the requirements of an attribute have been fully or largely met for the method's intended scope of application. An intermediate rating (medium blue) indicates that a method meets some, but not all of the requirements of the attribute. A low rating (light blue) indicates that the requirements of the attribute are not met or that no evidence is available in relation to the attribute for the method.

Construct validity	Essential	**Attribute 1** **Availability of information relating to the technical basis of the method** Information is provided on the technical basis of the method, in terms of its scientific underpinnings and data, in order to allow a judgement on the validity of the method to be made.		ATHEANA
		Comprehensive information on the technical basis and/or data underpinning the method is available and its application is discussed as part of the documentation of the method.	X	Justification Extensive technical basis (behavioural science, accident experience in NPPs and other industries, other disciplines) although the connection between this technical basis and the method steps itself is complex. ATHEANA provides a systematic, structured means to identify EFCs (Error Forcing Context search process). ATHEANA does not provide data. Technical Basis and Implementation Guideline (NUREG-1624, Rev. 1, 2000) and User's Guide (NUREG-1880, 2007)
		The method provides references that allow the information forming the technical basis and/or the data underpinning it to be obtained.		
		The method does not provide sufficient information to allow its technical basis and underpinning data to be accessed for review.		

Construct validity	Essential	**Attribute 2** **The technical basis of the method (Theory)** The technical basis of the method is based upon, and does not contradict, a relevant body of scientific knowledge		ATHEANA
		The method operationalises a relevant model of human performance or system safety which has scientific acceptance.	X	Justification The method is based on a classical information-processing model (four stages of cognition) monitoring/detection, situation assessment, response planning and response execution. ATHEANA's search process (identification of operational stories, combining Error Forcing Contexts (EFCs) and unsafe actions) is centred on a broad set of cognitive failure mechanisms.
		Elements of the method are inconsistent with an accepted scientific model of human performance or system safety.		

Construct validity	Essential	**Attribute 3** **The technical basis of the method (Data)** Where the technical basis of the method is based on a dataset, the source of the data/information and its relevance for application in the nuclear industry should be demonstrated.		ATHEANA
		The data underlying the method are largely based on observations of actual or simulated task performance in nuclear industry tasks.		Justification The ATHEANA method is based on a set of nuclear power plant event analyses. Failure probabilities are ultimately obtained from the expert judgement of the analysis team, during the application of the method.
		The data underlying the method are based on expert judgement or observations of human performance for relevant tasks in a domain that is closely related to the nuclear industry e.g. other high hazard industries.	X	
		The data underlying the method are taken from tasks that are not related or relevant to nuclear industry tasks.		

Construct validity	Highly desirable	**Attribute 4** **Internal consistency of the method** The method demonstrates internal consistency between the technical basis, the error definition, the PSFs and the qualitative and quantitative method steps		ATHEANA
		The qualitative and quantitative component parts of the method are theoretically compatible and form a coherent consistent whole.	X	Justification ATHEANA has been developed as an integral method addressing qualitative analysis, identification and modelling of the error forcing contexts/failure scenarios contributing to an HFE, and quantification of the failure scenarios. Note that the quantitative method steps provide guidance and structure for an expert elicitation and do not reference a dataset. In quantifying the HFE, the failure scenarios, which are identified in qualitative analysis and modelling, are quantified directly.
		There are theoretical inconsistencies between the qualitative and/or quantitative components of the method.		

Content Validity	Highly desirable	**Attribute 5** **Qualitative assessment** It is recognised good practice that HRA quantification is supported by qualitative analysis to develop an understanding of operator performance within the scenario that is being assessed. This attribute considers the extent to which the qualitative analysis stages of the HRA (e.g. task analysis and error identification) is directed or prescribed by the HRA method, beyond providing a set of performance shaping factors to be considered.		ATHEANA
		The method contains or prescribes a process for conducting qualitative assessment**.**	X	Justification ATHEANA provides detailed guidance for conducting a qualitative assessment of the HFE. Additionally, the application of the method inherently requires that this qualitative assessment be performed.
		The method includes a general statement indicating that a qualitative assessment should be provided, e.g. by referring to the use of task analysis**.**		
		The method does not make any reference to qualitative analysis**.**		

Content validity	Essential	**Attribute 6** **Factors influencing human reliability considered by the method** The method should be quantitatively sensitive to a majority of accepted factors* (PSFs) that influence human reliability. *: There are pre-defined lists of PSFs/EPCs available throughout the literature and within HRA methods (typically 1st generation methods). This attribute does not seek to define such a list, so as to accommodate developments in human performance, system safety and accident analysis. The evaluation teams are asked to use professional judgment when considering this attribute. It is not expected that methods will be assigned a high rating (dark blue) if only a small number of factors are accommodated, but we do not prescribe an absolute number of factors that are required.		ATHEANA
		Sub-scale 1 **Adequacy of PSFs**.		ATHEANA
		The method requires qualitative assessment of a majority of accepted factors that affect human reliability (PSFs).	X	Justification (Reviewers should identify the PSFs that are included in the method and judge the adequacy of this set for the intended application) At least the 16 PSFs listed in Section 5.2 of the ATHEANA User's Guide are expected to be addressed. This set corresponds to the set of PSFs in NUREG-1792, "Good Practices for Implementing Human Reliability Analysis (HRA)".
		The method does not consider a majority set of factors that affect human reliability.		

Content validity	Essential	**Sub-scale 2** **Quantitative sensitivity**		ATHEANA
		The method is quantitatively sensitive to the effect of each individual PSF considered qualitatively.		Justification The PSFs are reflected in the EFCs identified by the HRA analysis team, which are then quantified. The HEP is thereby sensitive to the set of PSFs pertinent to the EFC as a whole. It should be noted that an HFE may be modelled with several applicable EFCs, the contributions of which may be added.
		The method is not quantitatively sensitive to individual PSFs, but makes a single adjustment to the HEP based on the contribution of the overall combination of the PSFs considered.	X	
		The method is not quantitatively sensitive to PSFs.		
		Sub-scale 3 **Interaction between factors** Typically HRA methods adopt a linear multiplicative combination of PSFs. It is recognised that some PSFs may interact in other ways, e.g. a step change in the effect of one PSF once a threshold has been reached on a second PSF, or where the effect of the combination of two PSFs is far greater than multiplicative relationship would predict or where one PSF has a triggering effect on other PSFs in a causal chain.		ATHEANA
		Interactions between PSFs are accounted for on the basis of knowledge of the relationship between specific PSFs.	X	Justification The EFCs and quantification focus on the failure mechanism/narrative resulting from the qualitative interaction of the PSFs.
		Combinations of PSF effects are accounted for using a simple linear model.		
		Interactions between or combination of PSF effects are not considered by the method.		

Content validity	Essential	**Attribute 7** **Consideration of human error dependency** Modelling should include consideration of human error dependencies or common cause failures.		ATHEANA
		The method provides a procedure for identifying potential sources of dependence among Human Failure Events (HFEs) and/or sub-tasks of an HFE, and provides a method to derive conditional HEPs based on the systematic assessment of these sources of dependence.	X	Justification ATHEANA addresses a portion of the dependence issue through the development of the contexts, a part of which involves recognition of where there may be dependencies such as among crew members, because of the closeness of time, because of similar conditions, because of the effect of earlier events in an accident sequence, and so forth. There is special attention paid to the possible dependencies between an initial failure and recovering from the initial failure in Step 8 of the process where the recovery potential is considered before estimating the overall HEP for the HFE of interest. Further, if in the development of a scenario context it is recognized that multiple human actions of interest are involved in the scenario, quantification can be performed in ATHEANA with explicit consideration of possible dependencies among the human actions. This is accounted for during the expert elicitation with each expert deciding the quantitative effects of the identified dependencies. Nonetheless, accounting for dependencies among multiple HFEs appearing in the same scenario/sequence that have not been already addressed is still subject to analyst recognition during the PRA process and subsequent quantification accordingly. In other words, analysts need to review where quantified events get included in the PRA models to be sure that the appropriate dependencies were considered.
		The method identifies potential sources of dependence, but does not provide a process for linking these sources of dependence to a quantified model for deriving conditional HEPs.		
		The method does not address dependencies and common cause mechanisms among tasks and sub-tasks.		

Content validity	Essential	**Attribute 8** **Consideration of deviations and progressions in accident sequences** The method should provide a capability to accommodate: • Deviations from nominal accident scenarios due to: (A) Plant conditions: 1. Aleatory factors, such as sizes and locations of equipment failures and time sequences. 2. Complicating factors, such as coincident failures in control, instrumentation and support systems not normally modelled explicitly in PSA models. (B) Human failure scenarios due to crew organisational & operating practices that introduce opportunities to create new failure mechanisms. • Fault progressions including consequential faults and accident sequences encompassing Level 1 and Level 2 PSA which may involve extended time sequences and degraded operating environments should also be accommodated.	ATHEANA

Content validity	Essential	**Sub-scale 1** **Deviations**		ATHEANA
		The method provides for the qualitative and quantitative assessment of all the types of deviations in accident scenarios.	X	Justification "Deviations" are one of the PSFs considered in Step 5.2 (NUREG-1880). The guidance is a reminder to analysts to consider such variability and its impacts. The identification of significant deviations (deviations to which the crew response would be strongly sensitive) results in the quantification of the various cases.
		The method provides for the qualitative assessment of human error during fault progressions, but does not provide for the derivation of HEPs in support of this assessment.		
		The method does not provide a means to deal with deviations in accident scenarios		
		Sub-scale 2 **Fault progression**.		ATHEANA
		The method provides for the qualitative and quantitative assessment of human errors during fault progressions including level 1 to level 2 PSA fault progressions	X	Justification The conditions that could occur or would need to be represented for the analysis of HFEs in Level 2 PSA can be represented with ATHEANA's notion of EFCs.
		The method provides for the qualitative assessment of human error during fault progressions, but does not provide for the derivation of HEPs in support of this assessment.		
		The method does not provide for the qualitative and quantitative assessment of human errors during fault progressions.		

Content validity	Highly desirable	**Attribute 9** **Consideration of cognitive error** The method should be sensitive to the factors that influence the diagnosis and decision making component of the response to an initiating event.		ATHEANA
		The method estimates the probability of cognitive error based on the assessment of a set of factors that are known to affect diagnosis and decision making performance	X	Justification The method's development was largely based on the understanding of cognitive processes as described by Reason in "Human Error" and by advisors to the project such as Emilie Roth. As such, the identification of opportunities for mistakes is consistent with cognitive error models. The quantification process is primarily concerned with identifying the probabilities of such situations rather than trying to model HEPs separately. Recent applications have sought to separate these factors however.
		The method uses a simple model such as a time reliability curve as the primary factor for estimating the probability of cognitive error**.**		
		The method provides no way of estimating the likelihood of cognitive error**.**		

Content validity	Highly desirable	**Attribute 10** **Consideration of statistical uncertainty** The method should provide for statistical uncertainty analysis of derived human error probabilities.		ATHEANA
		The method derives uncertainty parameters from experience (either in-plant or from relevant simulator trials).	X	Justification ATHEANA's uncertainty results are obtained by considering the aleatory aspects of the scenario explicitly, eliciting HEP distributions from individual assessors, and then building a distribution representing the consensus of the assessors. ATHEANA guides a structured elicitation process and identifies sources of potential uncertainty to be considered to derive an uncertainty distribution for the HFE.
		The method provides generic uncertainty parameters, e.g. standardised error factors.		
		The method provides no uncertainty parameters.		

Content validity	Desirable	**Attribute 11** **Consideration of organisational issues** The method should consider the impact of organisational issues including safety-culture factors (attitudes and behaviours), and organisational process factors (e.g. command and control structures, conflicts of interest, communication and decision making protocols on human reliability).		ATHEANA
		Sub-scale 1 **Safety-culture factors** (attitudes and behaviours).		ATHEANA
		The method provides an adequate quantitative method to adjust HEPs based on an assessment of safety culture/safety climate**.**	X	Justification Safety-culture issues may be identified by the ATHEANA analysis team and incorporated into the EFC and its quantification. In ATHEANA, these factors are not quantified by modifying a base HEP value to account for safety-culture factors.
		The method provides a qualitative means to assess safety culture/safety climate, but does not include a process to modify HEPs based on the assessment**.**		
		The method does not take into account safety culture factors**.**		
		Sub-scale 2 **Process factors** (e.g. command and control structures, communication and decision making protocols on human reliability).		ATHEANA
		The method provides a quantitative method to assess process factors	X	Justification Process factors may be identified by the ATHEANA analysis team and incorporated into the EFC and its quantification. In ATHEANA, these factors are not quantified by modifying a base HEP value to account for these factors.
		The method provides a qualitative means to assess process factors, but does not include a process to modify HEPs based on the assessment**.**		
		The method does not take into account process factors**.**		

Empirical validity	Essential/Desirable	**Attribute 12** **Empirical validity** The method should demonstrate evidence of empirical validation exercises, peer review processes or community acceptance based on application and maturity.		ATHEANA
		Sub-scale1 **Statistical evidence**		ATHEANA
		The HEP estimates derived by the method have been shown to demonstrate good agreement with plant and /or simulator data for comparable tasks**.**		Justification ATHEANA was evaluated in the International and US HRA Empirical Studies; in these studies, the HEPs estimated by the analysis teams were assessed against a set of HEP confidence bounds for the HFEs of interest, which were derived from the observed performances of crews on simulators. While the results were generally positive, it should be noted that the Empirical Studies are not quantitative validation studies.
		The HEP estimates derived by the method have been shown to demonstrate good agreement with HEP estimates produced by other HRA methods for the same or comparable tasks**.**		
		The method has failed to derive comparable HEP estimates in tests of empirical validity or has not been subject to such assessments**.**	X	
		Sub-scale 2 **Verification/Peer review**		ATHEANA
		The method has been subject to peer review by a team of recognised HRA experts, and the peer review comments have been incorporated to the development of the method**.**	X	Justification Rev. 1 of the Technical Basis and Implementation Guide was based on a peer review of NUREG-1624. Rev. 1 was subsequently peer-reviewed. Additionally, the User's Guide was developed as a response to application experience and peer review.
		The method has been subject to peer review by a single, recognised HRA expert, and the comments have been incorporated to the development of the method**.**		
		The method has not been subject to independent peer review or the method has not been updated in response to peer review comments.		

Empirical validity	Essential/Desirable	**Sub-scale 3** **Application/Maturity**		ATHEANA
		The method has been extensively applied, internationally, for five or more years**.**		Justification • Full-scope application is limited. ATHEANA was applied in three PSAs focused on the Pressurized Thermal Shock (PTS) issue (2004). • ATHEANA may be used for a subset of the HFEs, given its resource requirements. • No licensee PSAs have been based on ATHEANA. • ATHEANA has been applied in international and U.S. HRA empirical studies.
		The method has been applied to a limited number of HRAs**.**	X	
		The method has not yet been applied to a HRA**.**		

Reliability	Essential	**Attribute 13** **Computer models and software tools** If a method incorporates the use of a computer model or software tool to analyse a human action, A QA programme should be applied to ensure quality of the design and validity of the output.		ATHEANA
		A relevant, recognised/accepted international standard has been applied to the software design and verification of the computer based HRA method/tool**.**		Justification Not applicable. No computer model or software.
		The design of the computer based HRA method/tool is based upon a documented QA process, which includes software verification**.**		
		There is no evidence that the design of the computer based HRA method/tool is based on a structured and validated software development or QA method that includes software verification**.**		

Reliability	Highly desirable	**Attribute 14** **Reliability and traceability** The method should provide consistent qualitative and quantitative information for comparable scenarios within analysts and between analysts for similar scenarios. The method should also provide sufficient information to facilitate tracing estimates back to input assumptions.		ATHEANA
		Sub-scale 1 **Within analyst consistency/reliability**		ATHEANA
		A formal comparison, amenable to statistical analysis, has been undertaken to demonstrate that the same HRA analyst provides consistent answers for analyses made at different times for the same scenario**.**		Justification No such an analysis has been performed.
		An informal comparison has been undertaken, which suggests good within analyst agreement for analyses made at different times**.**		
		There is no information available to suggest good within analyst agreement for analyses made at different times**.**	X	

Reliability	**Highly desirable**	**Sub-scale 2** **Between analyst consistency/reliability**		ATHEANA
		A formal comparison, amenable to statistical analysis, has been undertaken to demonstrate that different HRA analysts provide consistent answers for the same scenario**.**		Justification Studies of between-analyst consistency/reliability for ATHEANA are limited. (This applies to other methods as well.) The U.S. HRA empirical study addressed the consistency/reliability of analyses performed by different analysis teams with ATHEANA method (as well as other analysis teams applying ASEP, CBDT+HCR/ORE, and SPAR-H). Two analysis teams assessed 4 HFEs with ATHEANA. In the case of ATHEANA (as well as of the other methods) there was limited agreement in the HEPs obtained for the HFEs examined in the study. However, a detailed comparison of the HRAs of the HFEs found significant differences in the qualitative findings used by the analysis teams to estimate the HEPs. Furthermore, there were also significant differences in the quantification approach applied by the two ATHEANA teams. Consequently, in spite of the limitations of the U.S. HRA Empirical Study, its results suggest that the between-analyst consistency is superficial. In conclusion, this information does not suggest good between-analyst agreement.
		An informal comparison has been undertaken, which suggests good between analyst agreement**.**		
		There is no information available to suggest good between analyst agreement**.**	X	
		Sub-scale 3 **Traceability**		ATHEANA
		The method provides a procedure to ensure easy, complete traceability of the estimates of human performance in the HRA, such that an independent reviewer could trace back HEPs to relevant assumptions, models and data cited in the method**.**		Justification ATHEANA requires a comprehensive documentation of the analysis to support the analysis steps and, in particular, the quantification by expert judgement. This contributes substantially to the traceability of the HRA. No formalised documentation structure is provided. Traceability is adequate in terms of understanding the EFCs identified and the quantification. Nevertheless, understanding why a given set of EFCs is appropriate to the exclusion of others is likely to be difficult. This is a characteristic of methods that depend on analysts to identify failure narratives.
		The HRA method itself does not provide a procedure for traceability, but there is sufficient information available about the method to facilitate traceability, and enable an independent reviewer to understand what was done**.**	X	
		There is insufficient information available to facilitate traceability**.**		

Usability	Highly desirable	**Attribute 15** **Definition of method scope** The scope of the method should be clearly defined.		ATHEANA
		The scope of the method is clearly defined in a user manual and/or technical basis document**.**	X	Justification The scope is defined in NUREG-1624 and NUREG-1880 (Technical Basis and User's Guide documents) and explicitly includes pre and post initiating event human errors, pre and post core damage actions and operating modes other than full power.
		The scope of the method is described vaguely and some analyst judgement is required to determine its applicability to a particular human action/error**.**		
		The scope of the method is not defined**.**		

Usability	Highly desirable	**Attribute 16** **Qualitative outputs** The method should produce qualitative outputs that are useful to inform human factors and safety management improvements at the plant.		ATHEANA
		The method generates qualitative information to inform improvements to reduce the potential for human error that is explicitly related to each of the factors that are used in the method to derive an HEP.	X	Justification The failure narratives are by definition very specific in terms of the contributions to the failure of the HFE. These should be useful.
		The method generates qualitative information to inform improvements to reduce the potential for human error, but this is not explicitly linked to each of the factors used in the derivation of HEPs.		
		The method does not generate qualitative information to inform improvements to reduce the potential for human error.		

Usability	Highly desirable	**Attribute 17** **Qualitative uncertainty and quantitative conservatism** Methods should be able to reflect uncertainties related to qualitative information via conservatisms in the quantification process.		ATHEANA
		The method provides a mathematical procedure for adjusting the conservatism of the HEPs derived as a function of the level of certainty in the qualitative information collected during the assessment.	X	Justification This issue is addressed through the use of the formal elicitation process for quantification. NUREG-1880 describes, the expert elicitation process for ATHEANA quantification that directly developed a probability distribution.
		The method provides a general caution on the need to adjust the conservatism of HEPs as a function of the level of certainty in the qualitative information collected, but does not provide a mathematical procedure for doing so.		
		The method does not address the issue of uncertainties in qualitative information and the impact of this on derived HEPs.		

Usability	Desirable	**Attribute 18** **Availability of user documentation** The method should be supported by a detailed user documentation e.g., manual or instructions, which describes how the method should be applied.		ATHEANA
		The method contains user documentation that provides a detailed step-by-step procedure for all steps in the derivation of an HEP.	X	Justification US NRC (2000). Technical Basis and Implementation Guidelines for A Technique for Human Event Analysis (ATHEANA), NUREG-1624, Rev. 1, U.S. Nuclear Regulatory Commission, Washington, DC, USA. US NRC (2007). ATHEANA User's Guide – Final Report, NUREG-1880, U.S. Nuclear Regulatory Commission, Washington DC, USA.
		The method contains user documentation that provides a high level description of how it is applied to derive HEPs, but not all elements of the method are detailed as step-by-step procedures.		
		The method provides only a high level description of its method of application and or data tables for the derivation of HEPs.		

Usability	Desirable	**Attribute 19** **Use of limiting values** The method should provide limiting values. (Relevant Good Practice documents discuss limiting values that are used in member countries).		ATHEANA
		The method provides limiting values and advice on their application.		Justification The documentation does not identify the issue of lower limits on HFE probabilities. The sanity check step (8.9) does not address lower limits.
		The method provides advice on the need to limit claims on human performance but does not provide specific limiting values.		
		The method does not consider the use of limiting values.	X	

Usability	Indifferent/Essential	**Attribute 20** **Resources** A comparative estimate of the resources (time, cost, utility demands, level of specialist training required, level and type of knowledge required) needed to apply the method in comparison with other HRA methods.		ATHEANA
		The estimated cost of and time required for applying the HRA method is less than or comparable with that of other HRA methods**.**		Justification The application of ATHEANA is likely to require more time and facility resources when compared to other HRA methods.
		The estimated cost of and time required for applying the HRA method is in excess of that required for application of other HRA methods**.**	X	

A2.4 Attribute Evaluations – MERMOS

Desirable Attributes of HRA – Methods Evaluation Scale – MERMOS

Instructions to assessors

Indicate your evaluation of how well a method meets the requirements of an attribute by placing a tick in the relevant coloured box.

Provide a succinct justification for the rating you have allocated in the text box labelled justification.

You should complete an evaluation for each attribute and each sub-scale of an attribute.

If an attribute is not relevant to the method you are evaluating then record this in the justification box with an explanation as to why the attribute is not relevant.

Your evaluation should be based on the application of the method exactly as it is described in the method's user documentation. If there are modifications to the process which improve its application, but which are not formally recorded in an update to the method's documentation, then these can be noted in the justification box as potential improvements, but they should not be considered as part of the evaluation.

Within the method evaluation scale, a high rating (dark blue) indicates that the requirements of an attribute have been fully or largely met for the method's intended scope of application. An intermediate rating (medium blue) indicates that a method meets some, but not all of the requirements of the attribute. A low rating (light blue) indicates that the requirements of the attribute are not met or that no evidence is available in relation to the attribute for the method

Construct validity	Essential	**Attribute 1** **Availability of information relating to the technical basis of the method** Information is provided on the technical basis of the method, in terms of its scientific underpinnings and data, in order to allow a judgement on the validity of the method to be made.		MERMOS
		Comprehensive information on the technical basis and/or data underpinning the method is available and its application is discussed as part of the documentation of the method.	X	Justification User's guide for EDF users: • HT-54/02/020/A décembre 2002 "Guide application de la méthode MERMOS d'évaluation probabiliste de la fiabilité humaine pour les EPS de référence" (C. BLIRANDO, H.PESME). The document contains a general presentation of the method (framework and principle) and a step by step guide, illustrated with examples. • HT-54/98/006/A "MERMOS: principes de la méthode N4" (C. Bieder, F. CARA) • HT-54/98/007/B octobre 2000 "MERMOS: justifications théoriques" (E. DESMARES, F. CARA) The document contains the theoretical justifications of MERMOS. Both documents are written in French and are proprietary documents owned by EDF Conference papers providing an overview of the method are publicly available. • PSAM 4 (C. Bieder, P. Le-Bot, J-L Bonnet, F. CARA) "MERMOS: EDF's new advanced HRA method" • PSAM 5 (C. Bieder, S. Vidal, P. Le-Bot) "Feedback from the actual implementation of the MERMOS method"
		The method provides references that allow the information forming the technical basis and/or the data underpinning it to be obtained.		
		The method does not provide sufficient information to allow its technical basis and underpinning data to be accessed for review.		

Construct validity	Essential	**Attribute 2** **The Technical basis of the method (Theory)** The technical basis of the method is based upon, and does not contradict, a relevant body of scientific knowledge		MERMOS
		The method operationalises a relevant model of human performance or system safety which has scientific acceptance.	X	Justification MERMOS method was developed by a multidisciplinary team including: reliability engineers, EOP's experts, HRA analysts and specialists in sociology. The study is based upon the knowledge of the accident dynamism and the EOPs and puts the human factor in the centre of the system. Human actions are considered as the result of the whole operational system with multiple interactions between the components (the crew, the organisation, the EOPs and the MMI) and the process. This system complies with 3 functions: Strategy, Action and Diagnosis in order to bring the reactor in a safe condition. The failure of one of these functions can lead to the failure of the mission (HFE). The state and the orientation of the operation system (CICA*) can lead to the failure of the mission if they are inappropriate and persist in time (c.f. HT-54/98/007/B). CICA: MERMOS concept The CICAs refer to dynamic modes of organisation within the emergency operation system that are basically positive but may prove negative in a very specific situation.
		Elements of the method are inconsistent with an accepted scientific model of human performance or system safety.		

Construct validity	Essential	**Attribute 3** **The technical basis of the method (Data)** Where the technical basis of the method is based on a dataset, the source of the data/information and its relevance for application in the nuclear industry should be demonstrated.		MERMOS
		The data underlying the method are largely based on observations of actual or simulated task performance in nuclear industry tasks.		Justification Description of scenarios relies on simulator experiments but data itself mostly refers to expert judgement. Quantification of the HFEs relies on expert judgement. A specific method, named RETADE*, has been developed for the purpose of MERMOS in order to collect and aggregate expert judgement. Thus the method contains a process for acquiring data but does not provide data. Quantification is done by 3 analysts: each probability (situation features, recovery, CICA**) is discussed in order to obtain a consensus. * RETADE = Recueil et traitement des avis d'expert (Collection and processing of expert judgement) ** CICA = Caractéristiques Importantes de la Conduite Accidentelle (can be explained by some features of the situation).
		The data underlying the method are based on expert judgement or observations of human performance for relevant tasks in a domain that is closely related to the nuclear industry e.g. other high hazard industries.	X	
		The data underlying the method are taken from tasks that are not related or relevant to nuclear industry tasks.		

Construct validity	Highly desirable	**Attribute 4** **Internal consistency of the method** The method demonstrates internal consistency between the technical basis, the error definition, the PSFs and the qualitative and quantitative method steps		MERMOS
		The qualitative and quantitative component parts of the method are theoretically compatible and form a coherent consistent whole.	X	Justification MERMOS is made of two modules that have to be treated successively. The first one is dedicated to the identification and definition of the HFE: – what is required for the completion of the mission; – analysis of EOPs; – analysis of simulator studies; – description of the HFE in accordance to the template. The second one is the elaboration of failure scenarios through the analysis of 3 functions: strategy, diagnosis, action and their quantification. This way of proceeding insures the consistency of the analyses.
		There are theoretical inconsistencies between the qualitative and/or quantitative components of the method.		

Content Validity	Highly desirable	**Attribute 5** **Qualitative assessment** It is recognised good practice that HRA quantification is supported by qualitative analysis to develop an understanding of operator performance within the scenario that is being assessed. This attribute considers the extent to which the qualitative analysis stages of the HRA (e.g. task analysis and error identification) is directed or prescribed by the HRA method, beyond providing a set of performance shaping factors to be considered.		MERMOS
		The method contains or prescribes a process for conducting qualitative assessment**.**	X	Justification MERMOS analyses the failure of a human mission through failure scenarios. A failure scenario can be explained by certain modes of behaviour of the emergency operation system – the CICAs (Caractéristiques Importantes de la Conduite Accidentelle) – which can be explained by some features of the situation. MERMOS provides a description of the process by which failure scenarios can be envisioned.
		The method includes a general statement indicating that a qualitative assessment should be provided, e.g. by referring to the use of task analysis**.**		
		The method does not make any reference to qualitative analysis**.**		

Content validity	Essential	**Attribute 6** **Factors influencing human reliability considered by the method** The method should be quantitatively sensitive to a majority of accepted factors* (PSFs) that influence human reliability. *: There are pre-defined lists of PSFs/EPCs available throughout the literature and within HRA methods (typically 1st generation methods). This attribute does not seek to define such a list, so as to accommodate developments in human performance, system safety and accident analysis. The evaluation teams are asked to use professional judgment when considering this attribute. It is not expected that methods will be assigned a high rating (dark blue) if only a small number of factors are accommodated, but we do not prescribe an absolute number of factors that are required.		MERMOS
		Sub-scale 1 **Adequacy of PSFs**		MERMOS
		The method requires qualitative assessment of a majority of accepted factors that affect human reliability (PSFs).	X	Justification (Reviewers should identify the PSFs that are included in the method and judge the adequacy of this set for the intended application) One main step of MERMOS analysis consists in identifying the failure scenarios. They describe how the failure of the mission occurs. Features are of two types: structural (available time, EOPs, characteristics of the plant…) or contextual (information, stress, workload…). By this way, "classical" HRA PSFs are addressed.
		The method does not consider a majority set of factors that affect human reliability.		

Content validity	Essential	**Sub-scale 2** **Quantitative sensitivity**		MERMOS
		The method is quantitatively sensitive to the effect of each individual PSF considered qualitatively.	X	Justification MERMOS is sensitive to variation in the context that is part of the failure scenario used in MERMOS. The context contains the specifics of the PSFs being considered in the particular scenario. By varying the definition of the context and requantifying, the effects of the changes in the specific PSFs are incorporated in the results.
		The method is not quantitatively sensitive to individual PSFs, but makes a single adjustment to the HEP based on the contribution of the overall combination of the PSFs considered.		
		The method is not quantitatively sensitive to PSFs.		
		Sub-scale 3 **Interaction between factors** Typically HRA methods adopt a linear multiplicative combination of PSFs. It is recognised that some PSFs may interact in other ways, e.g. a step change in the effect of one PSF once a threshold has been reached on a second PSF, or where the effect of the combination of two PSFs is far greater than multiplicative relationship would predict or where one PSF has a triggering effect on other PSFs in a causal chain.		MERMOS
		Interactions between PSFs are accounted for on the basis of knowledge of the relationship between specific PSFs.	X	Justification The quantification of HFEs in MERMOS is based on the identification of one or more failure scenarios. Each failure scenario represents a specific configuration of how the characteristics of the PSA scenario (including the presentation of plant indications), of the procedures and training, and of the operating crews interact to result in the failure of the scenario. Consequently, while MERMOS is not a PSF-based method, the analytical process of identifying and quantifying failure scenarios explicitly addresses how the factors interact in a given scenario to result in a failure of the HFE.
		Combinations of PSF effects are accounted for using a simple linear model.		
		Interactions between or combination of PSF effects are not considered by the method.		

Content validity	Essential	**Attribute 7** **Consideration of human error dependency** Modelling should include consideration of human error dependencies or common cause failures.		MERMOS
		The method provides a procedure for identifying potential sources of dependence among Human Failure Events (HFEs) and/or sub-tasks of an HFE, and provides a method to derive conditional HEPs based on the systematic assessment of these sources of dependence.	X	Justification In case of consecutive HFEs, MERMOS advices to reword the second HFE in order that the failure of the first HFE becomes evident. The failure context is taken into account for the analysis of the second HFE (features of the scenario).
		The method identifies potential sources of dependence, but does not provide a process for linking these sources of dependence to a quantified model for deriving conditional HEPs.		
		The method does not address dependencies and common cause mechanisms among tasks and sub-tasks.		

Content validity	Essential	**Attribute 8** **Consideration of deviations and progressions in accident sequences** The method should provide a capability to accommodate: • Deviations from nominal accident scenarios due to: (A) Plant conditions: 1. Aleatory factors, such as sizes and locations of equipment failures and time sequences. 2. Complicating factors, such as coincident failures in control, instrumentation and support systems not normally modelled explicitly in PSA models. (B) Human failure scenarios due to crew organisational & operating practices that introduce opportunities to create new failure mechanisms. • Fault progressions including consequential faults and accident sequences encompassing Level 1 and Level 2 PSA which may involve extended time sequences and degraded operating environments should also be accommodated.		MERMOS
		Sub-scale 1 **Deviations**		
		The method provides for the qualitative and quantitative assessment of all the types of deviations in accident scenarios.	X	Justification MERMOS consist in searching failure scenarios leading to the failure of the human mission. Failure modes of the 3 functions "Strategy/Action/Diagnosis" are systematically analysed. The purpose is to imagine as many failure scenarios as possible by bridging the gap between theoretical concepts and real data. Failure scenarios can be predicted through a deductive or inductive approach starting from site or simulator feedback data that are available to him. The situation includes features of the process, of the crew, of the procedures through time.
		The method provides for the qualitative assessment of human error during fault progressions, but does not provide for the derivation of HEPs in support of this assessment.		
		The method does not provide a means to deal with deviations in accident scenarios		

Content validity	Essential	**Sub-scale 2** **Fault progression**.		MERMOS
		The method provides for the qualitative and quantitative assessment of human errors during fault progressions including level 1 to level 2 PSA fault progressions	X	Justification MERMOS has been employed for a try of modelling HFEs for level 2 PSA. It resulted to a derived method with 3 levels (only level 3 corresponds to the same methodology as MERMOS): - Level 1: Fixed values for the HFE (conservative expert judgement). - Level 2: Fixed values for the failure of function "Strategy/Action/Diagnosis"; a function "prognostic" was added; (conservative expert judgement). - Level 3: Scenarios are developed for the failure of function "Strategy/Action/ Diagnosis/Prognostic".
		The method provides for the qualitative assessment of human error during fault progressions, but does not provide for the derivation of HEPs in support of this assessment.		
		The method does not provide for the qualitative and quantitative assessment of human errors during fault progressions.		

Content validity	Highly desirable	**Attribute 9** **Consideration of cognitive error** The method should be sensitive to the factors that influence the diagnosis and decision making component of the response to an initiating event.		MERMOS
		The method estimates the probability of cognitive error based on the assessment of a set of factors that are known to affect diagnosis and decision making performance	X	Justification Cognitive errors are considered through the analysis of the failure mode of functions "Strategy" and "Diagnosis" : • 2 failure modes for the function Strategy: no strategy or wrong strategy; • 4 failure modes for the function Diagnosis: no diagnosis or wrong diagnosis for the state; no diagnosis or wrong diagnosis for the situation.
		The method uses a simple model such as a time reliability curve as the primary factor for estimating the probability of cognitive error**.**		
		The method provides no way of estimating the likelihood of cognitive error**.**		

Content validity	Highly desirable	**Attribute 10** **Consideration of statistical uncertainty** The method should provide for statistical uncertainty analysis of derived human error probabilities.		MERMOS
		The method derives uncertainty parameters from experience (either in-plant or from relevant simulator trials).		Justification A lognormal distribution is used by PSA analysts at EDF and MERMOS asks to apply an error factor of 10 for MERMOS HEPs: it is a conservative maximum value from international state of the art.
		The method provides generic uncertainty parameters, e.g. standardised error factors.	X	
		The method provides no uncertainty parameters.		

Content validity	**Desirable**	**Attribute 11** **Consideration of organisational issues** The method should consider the impact of organisational issues including safety-culture factors (attitudes and behaviours), and organisational process factors (e.g. command and control structures, conflicts of interest, communication and decision making protocols on human reliability).		MERMOS
		Sub-scale 1 **Safety-culture factors** (attitudes and behaviours).		
		The method provides an adequate quantitative method to adjust HEPs based on an assessment of safety culture/safety climate**.**	X	Justification The purpose of MERMOS method is to build failure scenarios that explain how the mission can be unsuccessful. A failure scenario can be explained by certain modes of organisation of the emergency operation system – the CICAs (important characteristics of emergency operation) – which can be explained by some features of the situation. A poor safety culture can be revealed in some features of the situation.
		The method provides a qualitative means to assess safety culture/safety climate, but does not include a process to modify HEPs based on the assessment**.**		
		The method does not take into account safety culture factors**.**		

Content validity	Desirable	**Sub-scale 2** **Process factors** (e.g. command and control structures, communication and decision making protocols on human reliability).		MERMOS
		The method provides a quantitative method to assess process factors	X	Justification MERMOS HRA method has the particularity to examine the human reliability from a unit larger than the crew. The purpose is to analyse the failure of a system taking into account the interactions between human components, hardware and the organisational environment in which all this system evolves and no longer to represent a linear pattern of the type: operator error, recovery by the shift supervisor or the safety engineer. The MERMOS analysis seeks to represent the coupling which exists between the crew, the Emergency Operating Procedures and the Man-Machine Interface which can lead to a poor outcome vis-à-vis the requirements for safety. As such the analysis takes into account the processes involved in managing safety by this plant system, which includes the process factors.
		The method provides a qualitative means to assess process factors, but does not include a process to modify HEPs based on the assessment**.**		
		The method does not take into account process factors**.**		
Empirical validity	Essential/Desirable	**Attribute 12 – Empirical validity** The method should demonstrate evidence of empirical validation exercises, peer review processes or community acceptance based on application and maturity.		MERMOS
		Sub-scale1 **Statistical evidence**		MERMOS
		The HEP estimates derived by the method have been shown to demonstrate good agreement with plant and /or simulator data for comparable tasks**.**		Justification MERMOS is implemented only at EDF. MERMOS participated to the international empirical HRA benchmark study (hosted in Halden) with rather good results. However due to the non-statistical treatment of the data generated by the international empirical study, it is not considered to provide evidence in relation to this attribute in this study.
		The HEP estimates derived by the method have been shown to demonstrate good agreement with HEP estimates produced by other HRA methods for the same or comparable tasks**.**		
		The method has failed to derive comparable HEP estimates in tests of empirical validity or has not been subject to such assessments**.**	X	

Empirical validity	Essential/Desirable	**Sub-scale 2** **Verification/Peer review**		MERMOS
		The method has been subject to peer review by a team of recognised HRA experts, and the peer review comments have been incorporated to the development of the method**.**		Justification An initial external review was done by SAIC and modifications were made before publishing the method. Several external assessments of the methods have been done, for example by HSE (RR679 Research report).
		The method has been subject to peer review by a single, recognised HRA expert, and the comments have been incorporated to the development of the method**.**	X	
		The method has not been subject to independent peer review or the method has not been updated in response to peer review comments.		
		Sub-scale 3 **Application/Maturity**		MERMOS
		The method has been extensively applied, internationally, for five or more years**.**		Justification MERMOS is applied to HRA studies at EDF since the years 2000. It has been applied to PSA studies for several series (900 MWe, 1 300 MWe and N4). Implementation is limited to EDF.
		The method has been applied to a limited number of HRAs**.**	X	
		The method has not yet been applied to a HRA**.**		

Reliability	Essential	**Attribute 13** **Computer models and software tools** If a method incorporates the use of a computer model or software tool to analyse a human action, A QA programme should be applied to ensure quality of the design and validity of the output.		MERMOS
		A relevant, recognised/accepted international standard has been applied to the software design and verification of the computer based HRA method/tool**.**		Justification N/A. The method does not use any software tool.
		The design of the computer based HRA method/tool is based upon a documented QA process, which includes software verification**.**		
		There is no evidence that the design of the computer based HRA method/tool is based on a structured and validated software development or QA method that includes software verification**.**		

Reliability	Highly desirable	**Attribute 14** **Reliability and traceability** The method should provide consistent qualitative and quantitative information for comparable scenarios within analysts and between analysts for similar scenarios. The method should also provide sufficient information to facilitate tracing estimates back to input assumptions.		MERMOS
		Sub-scale 1 **Within analyst consistency/reliability**		MERMOS
		A formal comparison, amenable to statistical analysis, has been undertaken to demonstrate that the same HRA analyst provides consistent answers for analyses made at different times for the same scenario**.**		Justification MERMOS has not been subject to any tests of within user reliability.
		An informal comparison has been undertaken, which suggests good within analyst agreement for analyses made at different times**.**		

Reliability	Highly desirable	There is no information available to suggest good within analyst agreement for analyses made at different times**.**	X	
		Sub-scale 2 **Between analyst consistency/reliability**		MERMOS
		A formal comparison, amenable to statistical analysis, has been undertaken to demonstrate that different HRA analysts provide consistent answers for the same scenario**.**		Justification MERMOS has not been subject to any tests of between user reliability. Quantification is done by a team of 3 analysts: each probability (situation features, recovery, CICA*) is discussed in order to obtain a consensus.
		An informal comparison has been undertaken, which suggests good between analyst agreement**.**		
		There is no information available to suggest good between analyst agreement**.**	X	
		Sub-scale 3 **Traceability**		MERMOS
		The method provides a procedure to ensure easy, complete traceability of the estimates of human performance in the HRA, such that an independent reviewer could trace back HEPs to relevant assumptions, models and data cited in the method**.**	X	Justification All along the evaluation, the analyst has to fill in the HFE form. Following information can be found: - Identification of the mission: name/code/initiating event/reactor state/EOPs. - Description of the mission: success criteria/reference marks/strategy of the mission/description of the action/parameters for diagnosis/the path of the operators through the EOPs. - A detailed description of the failure scenarios. - The probability of the HFE including the quantification of the parameters of all the scenarios. The HFE form represents about 10 to 20 pages.
		The HRA method itself does not provide a procedure for traceability, but there is sufficient information available about the method to facilitate traceability, and enable an independent reviewer to understand what was done**.**		
		There is insufficient information available to facilitate traceability**.**		

Usability	Highly desirable	**Attribute 15** **Definition of method scope** The scope of the method should be clearly defined.		MERMOS
		The scope of the method is clearly defined in a user manual and/or technical basis document**.**	X	Justification MERMOS was originally developed for the assessment of post-initiating event human errors, recent developments have extended the method to also consider pre-initiating human errors (MERMOS-A). MERMOS A has not been considered in this evaluation.
		The scope of the method is described vaguely and some analyst judgement is required to determine its applicability to a particular human action/error**.**		
		The scope of the method is not defined**.**		

Usability	Highly desirable	**Attribute 16** **Qualitative outputs** The method should produce qualitative outputs that are useful to inform human factors and safety management improvements at the plant		MERMOS
		The method generates qualitative information to inform improvements to reduce the potential for human error that is explicitly related to each of the factors that are used in the method to derive an HEP.	X	Justification MERMOS analyses the failure of a human mission through the dysfunctioning of 3 functions: Strategy/Action/Diagnosis (SAD). For each function, failure scenarios show how the failure can occur. A failure scenario can be explained by certain modes of organisation of the emergency operation system –the CICAs (Caractéristiques Importantes de la Conduite Accidentelle)- which can be explained by some features of the situation. This analysis can clearly by used to find ways of improvement.
		The method generates qualitative information to inform improvements to reduce the potential for human error, but this is not explicitly linked to each of the factors used in the derivation of HEPs.		
		The method does not generate qualitative information to inform improvements to reduce the potential for human error.		

Usability	Highly desirable	**Attribute 17** **Qualitative uncertainty and quantitative conservatism** Methods should be able to reflect uncertainties related to qualitative information via conservatisms in the quantification process.		MERMOS
		The method provides a mathematical procedure for adjusting the conservatism of the HEPs derived as a function of the level of certainty in the qualitative information collected during the assessment.		Justification MERMOS user's guide asks to refer to the report "Good practices on uncertainty for reference PSAs". This report is restricted to EDF.
		The method provides a general caution on the need to adjust the conservatism of HEPs as a function of the level of certainty in the qualitative information collected, but does not provide a mathematical procedure for doing so.	X	
		The method does not address the issue of uncertainties in qualitative information and the impact of this on derived HEPs.		
Usability	Desirable	**Attribute 18 – Availability of user documentation** The method should be supported by a detailed user documentation e.g., manual or instructions, which describes how the method should be applied.		MERMOS
		The method contains user documentation that provides a detailed step-by-step procedure for all steps in the derivation of an HEP.	X	Justification A user's manual is available for EDF users. It is made of two parts: one part is dedicated to the analysis of the HFE, the second one to the modelling. Each step is illustrated with examples. Moreover, EDF offers internal training sessions to analysts who implement the MERMOS method.
		The method contains user documentation that provides a high level description of how it is applied to derive HEPs, but not all elements of the method are detailed as step-by-step procedures.		
		The method provides only a high level description of its method of application and or data tables for the derivation of HEPs.		

Usability	Desirable	**Attribute 19** **Use of limiting values** The method should provide limiting values*. (Relevant Good Practice documents discuss limiting values that are used in member countries).		MERMOS
		The method provides limiting values and advice on their application.	X	Justification In MERMOS, for each HFE, a residual risk probability is calculated (from 10^{-4} to 10^{-5}). It has to be added to the probability of the failure scenarios. In consequence the minimum probability for an HEF is 10^{-5}.
		The method provides advice on the need to limit claims on human performance but does not provide specific limiting values.		
		The method does not consider the use of limiting values.		

Usability	Indifferent/Essential	**Attribute 20** **Resources** A comparative estimate of the resources (time, cost, utility demands, level of specialist training required, level and type of knowledge required) needed to apply the method in comparison with other HRA methods.		MERMOS
		The estimated cost of and time required for applying the HRA method is less than or comparable with that of other HRA methods**.**		Justification Full application of the original MERMOS approach which required conduct of simulator exercises involving station operating personnel is recognised to require significant resources. However, the more recently developed MERMOS C approach, which uses a catalogue of extant HFE assessments as the basis of the assessment, can reduce the resources required to complete the analysis such that the resources required are comparable with other HRA methods.
		The estimated cost of and time required for applying the HRA method is in excess of that required for application of other HRA methods**.**	X	

A2.5 Attribute Evaluations – NARA

Desirable Attributes of HRA – Methods Evaluation Scale – NARA

Instructions to assessors

Indicate your evaluation of how well a method meets the requirements of an attribute by placing a tick in the relevant coloured box.

Provide a succinct justification for the rating you have allocated in the text box labelled justification.

You should complete an evaluation for each attribute and each sub-scale of an attribute.

If an attribute is not relevant to the method you are evaluating then record this in the justification box with an explanation as to why the attribute is not relevant.

Your evaluation should be based on the application of the method exactly as it is described in the method's user documentation. If there are modifications to the process which improve its application, but which are not formally recorded in an update to the method's documentation, then these can be noted in the justification box as potential improvements, but they should not be considered as part of the evaluation.

Within the method evaluation scale, a high rating (dark blue) indicates that the requirements of an attribute have been fully or largely met for the method's intended scope of application. An intermediate rating (medium blue) indicates that a method meets some, but not all of the requirements of the attribute. A low rating (light blue) indicates that the requirements of the attribute are not met or that no evidence is available in relation to the attribute for the method.

Construct validity	Essential	**Attribute 1** **Availability of information relating to the technical basis of the method** Information is provided on the technical basis of the method, in terms of its scientific underpinnings and data, in order to allow a judgement on the validity of the method to be made.		NARA
		Comprehensive information on the technical basis and/or data underpinning the method is available and its application is discussed as part of the documentation of the method.	X	Justification Comprehensive documentation (367 pages) on the NARA technique and its derivation is provided in the Technical Basis for NARA, a Method of Human Error Quantification, Issue 7, January 2012, Report CRA-BEGL-POW-J032. There is also a shorter User Manual. The Technical Basis contains all data used in derivation of the quantification aspects of the technique. The technical basis document also provides a discussion on the relationship between the NARA technique and human information processing models of human performance. The technical basis document details how the values (HEPs and EPC weights) are derived from data, the sources of all data being identified. The technical basis document is a proprietary document owned by EDF Nuclear Generation Limited and is not publically available.
		The method provides references that allow the information forming the technical basis and/or the data underpinning it to be obtained.		
		The method does not provide sufficient information to allow its technical basis and underpinning data to be accessed for review.		

Construct validity	Essential	**Attribute 2** **The Technical basis of the method (Theory)** The technical basis of the method is based upon, and does not contradict, a relevant body of scientific knowledge		NARA
		The method operationalises a relevant model of human performance or system safety which has scientific acceptance.	X	Justification NARA is not a direct operationalisation of a single model of human performance or system safety. The NARA Technical Basis document provides a discussion of the technical basis of the method demonstrating how it relates to three error-related modelling traditions in Human Factors & Performance: Information Processing, the Skill, Rule and Knowledge-Based Behaviour model, and Reason's "Slips, lapses and mistakes" model. The method therefore is not inconsistent with accepted scientific models. However, neither is it a direct operationalisation of relevant models.
		Elements of the method are inconsistent with an accepted scientific model of human performance or system safety.		

Construct validity	Essential	**Attribute 3** **The technical basis of the method (Data)** Where the technical basis of the method is based on a dataset, the source of the data/information and its relevance for application in the nuclear industry should be demonstrated.		NARA
		The data underlying the method are largely based on observations of actual or simulated task performance in nuclear industry tasks.	X	Justification The NARA Technical basis document identifies each data point used in the derivation of HEPs associated with each Generic Task Type (GTT) used in the method. Approximately 2/3 of these come from the nuclear industry with the remainder deriving from other industries. The technical basis document also identifies each of the data sources used in establishing the maximum affect associated with each Error Producing Condition (NARA term for PSF). The majority of data used to derive numeric values for this aspect of the method are from laboratory experiments often using simple tasks that form component parts of nuclear industry tasks.
		The data underlying the method are based on expert judgement or observations of human performance for relevant tasks in a domain that is closely related to the nuclear industry e.g. other high hazard industries.		
		The data underlying the method are taken from tasks that are not related or relevant to nuclear industry tasks.		

Construct validity	Highly desirable	**Attribute 4** **Internal consistency of the method** The method demonstrates internal consistency between the technical basis, the error definition, the PSFs and the qualitative and quantitative method steps**.**		NARA
		The qualitative and quantitative component parts of the method are theoretically compatible and form a coherent consistent whole.	X	Justification NARA demonstrates internal consistency between the quantification procedures and the theoretical basis which is largely founded on an information processing model. The quantification processes themselves are internally consistent with HEPs being assigned on the basis of generic task characteristics and these being modified on the basis of performance shaping factors including extended time factors.
		There are theoretical inconsistencies between the qualitative and/or quantitative components of the method.		

Content Validity	Highly desirable	**Attribute 5** **Qualitative assessment** It is recognised good practice that HRA quantification is supported by qualitative analysis to develop an understanding of operator performance within the scenario that is being assessed. This attribute considers the extent to which the qualitative analysis stages of the HRA (e.g. task analysis and error identification) is directed or prescribed by the HRA method, beyond providing a set of performance shaping factors to be considered.		NARA
		The method contains or prescribes a process for conducting qualitative assessment**.**		Justification The NARA user manual identifies that a task and error analysis should be conducted wherever possible to underpin the quantitative analysis provided by NARA. The Manual also identifies that such qualitative analysis is outside of the scope of the manual.
		The method includes a general statement indicating that a qualitative assessment should be provided, e.g. by referring to the use of task analysis**.**	X	
		The method does not make any reference to qualitative analysis**.**		

Content validity	Essential	**Attribute 6** **Factors influencing human reliability considered by the method** The method should be quantitatively sensitive to a majority of accepted factors* (PSFs) that influence human reliability. *: There are pre-defined lists of PSFs/EPCs available throughout the literature and within HRA methods (typically 1st generation methods). This attribute does not seek to define such a list, so as to accommodate developments in human performance, system safety and accident analysis. The evaluation teams are asked to use professional judgment when considering this attribute. It is not expected that methods will be assigned a high rating (dark blue) if only a small number of factors are accommodated, but we do not prescribe an absolute number of factors that are required.	NARA

Content validity	Essential	**Sub-scale 1** **Adequacy of PSFs**.		
		The method requires qualitative assessment of a majority of accepted factors that affect human reliability (PSFs).	X	Justification (Reviewers should identify the PSFs that are included in the method and judge the adequacy of this set for the intended application) NARA has 18 PSFs (called error Producing Conditions, EPCs) listed below that should be considered during application. 1. Poor, ambiguous or ill-matched system feedback. 2. Unfamiliarity. 3. A need to unlearn a technique and apply one which requires the application of an opposing philosophy. 4. Time pressure. 5. Low signal to noise ratio. 6. A conflict between immediate and long-term objectives. 7. No obvious means of reversing an unintended action. 8. A means of suppressing or over-riding information or features which is too easily accessible. 9. Operator inexperience. 10. Cognitive overload, particularly one caused by simultaneous presentation of non-redundant information. 11. No obvious way of keeping track of progress during an activity. 12. Shortfalls in the quality of information conveyed by procedures. 13. Difficulties caused by poor shift hand-over practices and/or team co-ordination problems or friction between team members. 14. An incentive to use other more dangerous procedures to achieve long-term objectives. 15. Poor environment. 16. High emotional stress and effects of ill health. 17. Low workforce morale or adverse organisational environment. 18. Operator under-load/boredom. The set of PSFs overlaps with those used in other HRA methods of this type and is consistent with relevant good practice as outlined e.g. in the USNRC Good Practices for implementing HRA guidance.
		The method does not consider a majority set of factors that affect human reliability.		

Content validity	Essential	**Sub-scale 2** **Quantitative sensitivity**.		NARA
		The method is quantitatively sensitive to the effect of each individual PSF considered qualitatively.	X	Justification Each PSF (EPC) has its own independent quantitative weighting (effect on performance reliability).
		The method is not quantitatively sensitive to individual PSFs, but makes a single adjustment to the HEP based on the contribution of the overall combination of the PSFs considered.		
		The method is not quantitatively sensitive to PSFs.		
		Sub-scale 3 **Interaction between factors** Typically HRA methods adopt a linear multiplicative combination of PSFs. It is recognised that some PSFs may interact in other ways, e.g. a step change in the effect of one PSF once a threshold has been reached on a second PSF, or where the effect of the combination of two PSFs is far greater than multiplicative relationship would predict or where one PSF has a triggering effect on other PSFs in a causal chain.		NARA
		Interactions between PSFs are accounted for on the basis of knowledge of the relationship between specific PSFs.		Justification No PSF (EPC) interaction effects are considered. NARA uses a simple linear model.
		Combinations of PSF effects are accounted for using a simple linear model.	X	
		Interactions between or combination of PSF effects are not considered by the method.		

Content validity	Essential	**Attribute 7** **Consideration of human error dependency** Modelling should include consideration of human error dependencies or common cause failures.		NARA
		The method provides a procedure for identifying potential sources of dependence among Human Failure Events (HFEs) and/or sub-tasks of an HFE, and provides a method to derive conditional HEPs based on the systematic assessment of these sources of dependence.		Justification The method does not provide a procedure for dealing with dependency. It does however provide guidance on methods that can be used to address the issue, but these are not an integral part of the method.
		The method identifies potential sources of dependence, but does not provide a process for linking these sources of dependence to a quantified model for deriving conditional HEPs.	X	
		The method does not address dependencies and common cause mechanisms among tasks and sub-tasks.		

Content validity	Essential	**Attribute 8** **Consideration of deviations and progressions in accident sequences** The method should provide a capability to accommodate: • Deviations from nominal accident scenarios due to: (A) Plant conditions: 1. Aleatory factors, such as sizes and locations of equipment failures and time sequences. 2. Complicating factors, such as coincident failures in control, instrumentation and support systems not normally modelled explicitly in PSA models. (B) Human failure scenarios due to crew organisational & operating practices that introduce opportunities to create new failure mechanisms. • Fault progressions including consequential faults and accident sequences encompassing Level 1 and Level 2 PSA which may involve extended time sequences and degraded operating environments should also be accommodated.	NARA

Content validity	Essential	**Sub-scale 1** **Deviations**		
		The method provides for the qualitative and quantitative assessment of all the types of deviations in accident scenarios.		Justification NARA does not provide the qualitative assessment tools required to model deviations in accident sequences.
		The method provides for the qualitative assessment of human error during fault progressions, but does not provide for the derivation of HEPs in support of this assessment.		
		The method does not provide a means to deal with deviations in accident scenarios	X	
		Sub-scale 2 **Fault progression**.		NARA
		The method provides for the qualitative and quantitative assessment of human errors during fault progressions including level 1 to level 2 PSA fault progressions		Justification NARA does not provide the qualitative assessment tools required to model progressions in accident sequences. NARA contains a method for dealing with extended time factors where these may have a positive impact on human performance, this aspect of the method may provide some benefit for considering fault progressions.
		The method provides for the qualitative assessment of human error during fault progressions, but does not provide for the derivation of HEPs in support of this assessment.		
		The method does not provide for the qualitative and quantitative assessment of human errors during fault progressions.	X	

Content validity	Highly desirable	**Attribute 9** **Consideration of cognitive error** The method should be sensitive to the factors that influence the diagnosis and decision making component of the response to an initiating event.		NARA
		The method estimates the probability of cognitive error based on the assessment of a set of factors that are known to affect diagnosis and decision making performance	X	Justification NARA contains three GTTs relevant to cognitive error which map onto Rasmussen's Skill, Rule, Knowledge framework. A number of EPCs that affect decision-making and diagnosis e.g. cognitive overload, low signal to noise ratio are considered by the method.
		The method uses a simple model such as a time reliability curve as the primary factor for estimating the probability of cognitive error**.**		
		The method provides no way of estimating the likelihood of cognitive error**.**		

Content validity	Highly desirable	**Attribute 10** **Consideration of statistical uncertainty** The method should provide for statistical uncertainty analysis of derived human error probabilities.		NARA
		The method derives uncertainty parameters from experience (either in-plant or from relevant simulator trials).		Justification The HEPs associated with GTTs have uncertainty bounds (5-95%) which are statistically-derived based on the number of data points (and their range) used to derive the GTT.
		The method provides generic uncertainty parameters, e.g. standardised error factors.	X	
		The method provides no uncertainty parameters.		

Content validity	Desirable	**Attribute 11** **Consideration of organisational issues** The method should consider the impact of organisational issues including safety-culture factors (attitudes and behaviours), and organisational process factors (e.g. command and control structures, conflicts of interest, communication and decision making protocols on human reliability).		NARA
		Sub-scale 1 **Safety-culture factors** (attitudes and behaviours).		NARA
		The method provides an adequate quantitative method to adjust HEPs based on an assessment of safety culture/safety climate**.**	X	Justification NARA considers a number of EPCs that relate to some aspects of safety culture e.g. a conflict between immediate and long term objectives, an incentive to use other, more dangerous procedures to achieve long-term objectives, low work force morale or adverse organisational environment. Whilst the EPCs may not address all of the components of safety culture, they provide some basic relevant factors to be addressed quantitatively.
		The method provides a qualitative means to assess safety culture/safety climate, but does not include a process to modify HEPs based on the assessment**.**		
		The method does not take into account safety culture factors**.**		
		Sub-scale 2 **Process factors** (e.g. command and control structures, communication and decision making protocols on human reliability).		NARA
		The method provides a quantitative method to assess process factors	X	Justification The method provides a number of EPCs that relate to organisational process factors these include: – Difficulties caused by poor shift hand-over practices and/or team co-ordination problems or friction between team members. – Incentives to use more dangerous procedures. – Low workforce morale or adverse organisational environment. However the whole set of organisational process factors are not considered to be addressed by the method.
		The method provides a qualitative means to assess process factors, but does not include a process to modify HEPs based on the assessment**.**		
		The method does not take into account process factors**.**		

Empirical validity	Essential/Desirable	**Attribute 12** **Empirical validity** The method should demonstrate evidence of empirical validation exercises, peer review processes or community acceptance based on application and maturity.		NARA
		Sub-scale1 **Statistical evidence**		NARA
		The HEP estimates derived by the method have been shown to demonstrate good agreement with plant and /or simulator data for comparable tasks**.**		Justification NARA has not been subjected to empirical validations.
		The HEP estimates derived by the method have been shown to demonstrate good agreement with HEP estimates produced by other HRA methods for the same or comparable tasks**.**		
		The method has failed to derive comparable HEP estimates in tests of empirical validity or has not been subject to such assessments**.**	X	
		Sub-scale 2 **Verification/Peer review**		NARA
		The method has been subject to peer review by a team of recognised HRA experts, and the peer review comments have been incorporated to the development of the method**.**	X	Justification NARA has been subject to two independent international Peer Reviews with six HRA experts who gave anonymous comments which the NARA development team had to respond to and resolve to the satisfaction of the experts and the regulator who commissioned the reviews. The reviews resulted in a number of changes to the method.
		The method has been subject to peer review by a single, recognised HRA expert, and the comments have been incorporated to the development of the method**.**		
		The method has not been subject to independent peer review or the method has not been updated in response to peer review comments.		

Empirical validity	Essential/Desirable	**Sub-scale 3** **Application/Maturity**		NARA
		The method has been extensively applied, internationally, for five or more years**.**		Justification NARA has been applied to only a limited number of HRAs in the United Kingdom having only recently replaced HEART as an identified method for conducting HRA within EDF NGL. NARA was used as a quantification tool in the United States in the Yucca Mountain HRA.
		The method has been applied to a limited number of HRAs**.**	X	
		The method has not yet been applied to a HRA**.**		

Reliability	Essential	**Attribute 13** **Computer models and software tools** If a method incorporates the use of a computer model or software tool to analyse a human action, A QA programme should be applied to ensure quality of the design and validity of the output.		NARA
		A relevant, recognised/accepted international standard has been applied to the software design and verification of the computer based HRA method/tool**.**		Justification Not Applicable. NARA has not been developed as a software tool.
		The design of the computer based HRA method/tool is based upon a documented QA process, which includes software verification**.**		
		There is no evidence that the design of the computer based HRA method/tool is based on a structured and validated software development or QA method that includes software verification**.**		

Reliability	Highly desirable	**Attribute 14** **Reliability and traceability** The method should provide consistent qualitative and quantitative information for comparable scenarios within analysts and between analysts for similar scenarios. The method should also provide sufficient information to facilitate tracing estimates back to input assumptions.		NARA
		Sub-scale 1 **Within analyst consistency/reliability**		
		A formal comparison, amenable to statistical analysis, has been undertaken to demonstrate that the same HRA analyst provides consistent answers for analyses made at different times for the same scenario**.**		Justification NARA has not been subject to any tests of within user reliability.
		An informal comparison has been undertaken, which suggests good within analyst agreement for analyses made at different times**.**		
		There is no information available to suggest good within analyst agreement for analyses made at different times**.**	X	
		Sub-scale 2 **Between analyst consistency/reliability**		NARA
		A formal comparison, amenable to statistical analysis, has been undertaken to demonstrate that different HRA analysts provide consistent answers for the same scenario**.**		Justification NARA has not been subject to any tests of between user reliability.
		An informal comparison has been undertaken, which suggests good between analyst agreement**.**		
		There is no information available to suggest good between analyst agreement**.**	X	

Reliability	Highly desirable	**Sub-scale 3** **Traceability**		NARA
		The method provides a procedure to ensure easy, complete traceability of the estimates of human performance in the HRA, such that an independent reviewer could trace back HEPs to relevant assumptions, models and data cited in the method**.**	X	Justification There is a strong emphasis on documentation and proforma for recording all assumptions, calculations, etc. This is dealt with in the User Manual and examples are given of how to document NARA usage. This is also heavily emphasised in the NARA training.
		The HRA method itself does not provide a procedure for traceability, but there is sufficient information available about the method to facilitate traceability, and enable an independent reviewer to understand what was done**.**		
		There is insufficient information available to facilitate traceability**.**		

Usability	Highly desirable	**Attribute 15** **Definition of method scope** The scope of the method should be clearly defined.		NARA
		The scope of the method is clearly defined in a user manual and/or technical basis document**.**		Justification The scope of the method is identified in the User Manual as to address the HEP assessment needs for a typical PSA, although what is meant by a typical PSA is not defined. The user manual also identifies that NARA is developed to consider fault sequences covering extended timescales and the need to cover dependencies. It is also identified that a module to cover errors of commission was produced but that this was considered tentative and as a result was not incorporated into the final technique. The user manual provides worked examples illustrating how NARA may be applied to quantification of pre-fault (maintenance) errors and post fault errors covering both diagnosis and action components including control room and on plant actions.
		The scope of the method is described vaguely and some analyst judgement is required to determine its applicability to a particular human action/error**.**	X	
		The scope of the method is not defined**.**		

Usability	Highly desirable	**Attribute 16** **Qualitative outputs** The method should produce qualitative outputs that are useful to inform human factors and safety management improvements at the plant**.**		NARA
		The method generates qualitative information to inform improvements to reduce the potential for human error that is explicitly related to each of the factors that are used in the method to derive an HEP.	X	Justification The output from a NARA analysis identifies the EPCs that have been used in deriving the HEP value. Information is available in the discussion of EPCs and their anchor values which can be used to derive recommendations for plant and operational improvements.
		The method generates qualitative information to inform improvements to reduce the potential for human error, but this is not explicitly linked to each of the factors used in the derivation of HEPs.		
		The method does not generate qualitative information to inform improvements to reduce the potential for human error.		

Usability	Highly desirable	**Attribute 17** **Qualitative uncertainty and quantitative conservatism** Methods should be able to reflect uncertainties related to qualitative information via conservatisms in the quantification process.		NARA
		The method provides a mathematical procedure for adjusting the conservatism of the HEPs derived as a function of the level of certainty in the qualitative information collected during the assessment.		Justification The User Manual does not explicitly address the issue of uncertainty related to qualitative information, however, NARA does provide a mechanism, the assessed proportion of affect, by which such uncertainties could be taken into account when deriving HEPs.
		The method provides a general caution on the need to adjust the conservatism of HEPs as a function of the level of certainty in the qualitative information collected, but does not provide a mathematical procedure for doing so.	X	
		The method does not address the issue of uncertainties in qualitative information and the impact of this on derived HEPs.		

Usability	Desirable	**Attribute 18** **Availability of user documentation** The method should be supported by a detailed user documentation e.g., manual or instructions, which describes how the method should be applied.		NARA
		The method contains user documentation that provides a detailed step-by-step procedure for all steps in the derivation of an HEP.	X	Justification NARA provides a User Manual which gives a detailed step-by-step procedure for all aspects of NARA application (GTT selection, EPC selection and anchoring, quantification, extended time factor consideration, and documenting the analysis.
		The method contains user documentation that provides a high level description of how it is applied to derive HEPs, but not all elements of the method are detailed as step-by-step procedures.		
		The method provides only a high level description of its method of application and or data tables for the derivation of HEPs.		

Usability	Desirable	**Attribute 19** **Use of limiting values** The method should provide limiting values. (Relevant Good Practice documents discuss limiting values that are used in member countries).		NARA
		The method provides limiting values and advice on their application.	X	Justification Human Performance Limiting Values are prescribed in the User Manual and their detailed consideration and application further explained in the Technical Basis document (Appendix J).
		The method provides advice on the need to limit claims on human performance but does not provide specific limiting values.		
		The method does not consider the use of limiting values.		

Usability	Indifferent/Essential	**Attribute 20** **Resources** A comparative estimate of the resources (time, cost, utility demands, level of specialist training required, level and type of knowledge required) needed to apply the method in comparison with other HRA methods.		NARA
		The estimated cost of and time required for applying the HRA method is less than or comparable with that of other HRA methods**.**	X	Justification The HEP quantification is quick relative to other techniques. As with any HRA technique, the real effort occurs in the qualitative analysis underpinning the HRA and the time required for this should be comparable with that for the application of other techniques. NARA currently has a mandatory 1.5 day training course to be completed by any assessor wishing to use the technique.
		The estimated cost of and time required for applying the HRA method is in excess of that required for application of other HRA methods**.**		

A2.6 **Attribute Evaluations – SPAR-H**

Desirable Attributes of HRA – Methods Evaluation Scale – SPAR-H

Instructions to assessors

Indicate your evaluation of how well a method meets the requirements of an attribute by placing a tick in the relevant coloured box.

Provide a succinct justification for the rating you have allocated in the text box labelled justification.

You should complete an evaluation for each attribute and each sub-scale of an attribute.

If an attribute is not relevant to the method you are evaluating then record this in the justification box with an explanation as to why the attribute is not relevant.

Your evaluation should be based on the application of the method exactly as it is described in the method's user documentation. If there are modifications to the process which improve its application, but which are not formally recorded in an update to the method's documentation, then these can be noted in the justification box as potential improvements, but they should not be considered as part of the evaluation.

Within the method evaluation scale, a high rating (dark blue) indicates that the requirements of an attribute have been fully or largely met for the method's intended scope of application. An intermediate rating (medium blue) indicates that a method meets some, but not all of the requirements of the attribute. A low rating (light blue) indicates that the requirements of the attribute are not met or that no evidence is available in relation to the attribute for the method.

Construct validity	Essential	**Attribute 1** **Availability of information relating to the technical basis of the method** Information is provided on the technical basis of the method, in terms of its scientific underpinnings and data, in order to allow a judgement on the validity of the method to be made.		SPAR-H
		Comprehensive information on the technical basis and/or data underpinning the method is available and its application is discussed as part of the documentation of the method.	X	Justification NUREG/CR-6883, The SPAR-H Human Reliability Analysis Method, is quite comprehensive including technical basis and application examples.
		The method provides references that allow the information forming the technical basis and/or the data underpinning it to be obtained.		
		The method does not provide sufficient information to allow its technical basis and underpinning data to be accessed for review.		

Construct validity	Essential	**Attribute 2** **The technical basis of the method (Theory)** The technical basis of the method is based upon, and does not contradict, a relevant body of scientific knowledge		SPAR-H
		The method operationalises a relevant model of human performance or system safety which has scientific acceptance.	X	Justification SPAR-H is based on a relevant model of human performance (the human information processing model) which has wide scientific acceptance (Baddeley, 1990; Sanders & McCormick, 1993).
		Elements of the method are inconsistent with an accepted scientific model of human performance or system safety.		

Construct validity	Essential	**Attribute 3** **The technical basis of the method (Data)** Where the technical basis of the method is based on a dataset, the source of the data/information and its relevance for application in the nuclear industry should be demonstrated.		SPAR-H
		The data underlying the method are largely based on observations of actual or simulated task performance in nuclear industry tasks.		Justification The user manual (NUREG/CR-6883) contains comparisons between a number of methods, both of nominal HEPs for task types and for PSFs. The background and history for the PSF weights are described in the paper: Boring, R.L., & Blackman, H.S. (2007). "The origins of the SPAR-H method's performance shaping factor multipliers". Official Proceedings of the Joint 8th IEEE Conference on Human Factors and Power Plants and the 13th Annual Workshop on Human Performance/Root Cause/Trending/Operating Experience/Self Assessment, 177-184. In this paper it is stated that the nominal HEPs are derived from THERP and WASH-1400, these have been kept constant in all the versions, 1994, 1999 and 2005. The original PSF weights were derived from THERP, and refinements in the newer updates were based on expert judgement and revision of other methods. So the comparisons to other methods are also a background for the values themselves.
		The data underlying the method are based on expert judgement or observations of human performance for relevant tasks in a domain that is closely related to the nuclear industry e.g. other high hazard industries.	X	
		The data underlying the method are taken from tasks that are not related or relevant to nuclear industry tasks.		

Construct validity	Highly desirable	**Attribute 4** **Internal consistency of the method** The method demonstrates internal consistency between the technical basis, the error definition, the PSFs and the qualitative and quantitative method steps		SPAR-H
		The qualitative and quantitative component parts of the method are theoretically compatible and form a coherent consistent whole.	X	Justification A qualitative analysis must be part of the basis for the judgements of the PSF weights. This is coherent, although there may be flaws in the basis for judging the PSF weights if a detailed enough qualitative analysis is not performed. SPAR-H assumes that both diagnosis and action (execution) aspects of performance contribute to the potential for error. This is consistent with an information processing view of cognition, which breaks performance down into Perception → Decision Making → Response Execution, in which response execution is at least partially independent of decision making. Research in cognitive science has demonstrated that contextual factors modify the potential for error in each of these stages. As in the psychological literature, environmental factors such as stress, workload, characteristics of the interface, etc., can have differential effects on these stages of performance. This is reflected in different values for performance shaping factors for different error types. SPAR-H multiplies the base rate of error by the factors assumed to be present in the context.
		There are theoretical inconsistencies between the qualitative and/or quantitative components of the method.		

Content Validity	Highly desirable	**Attribute 5** **Qualitative assessment** It is recognised good practice that HRA quantification is supported by qualitative analysis to develop an understanding of operator performance within the scenario that is being assessed. This attribute considers the extent to which the qualitative analysis stages of the HRA (e.g. task analysis and error identification) is directed or prescribed by the HRA method, beyond providing a set of performance shaping factors to be considered.	SPAR-H	
		The method contains or prescribes a process for conducting qualitative assessment**.**		Justification The SPAR-H method is a quantification method, and explicitly states so. For qualitative analysis, it refers to and summarizes ATHEANA's ten step search process.
		The method includes a general statement indicating that a qualitative assessment should be provided, e.g. by referring to the use of task analysis**.**	X	
		The method does not make any reference to qualitative analysis**.**		

Content validity	Essential	**Attribute 6** **Factors influencing human reliability considered by the method** The method should be quantitatively sensitive to a majority of accepted factors* (PSFs) that influence human reliability. *: There are pre-defined lists of PSFs/EPCs available throughout the literature and within HRA methods (typically 1st generation methods). This attribute does not seek to define such a list, so as to accommodate developments in human performance, system safety and accident analysis. The evaluation teams are asked to use professional judgment when considering this attribute. It is not expected that methods will be assigned a high rating (dark blue) if only a small number of factors are accommodated, but we do not prescribe an absolute number of factors that are required.	SPAR-H	
		Sub-scale 1 **Adequacy of PSFs**	SPAR-H	
		The method requires qualitative assessment of a majority of accepted factors that affect human reliability (PSFs).	X	Justification (Reviewers should identify the PSFs that are included in the method and judge the adequacy of this set for the intended application) The list of PSFs in SPAR-H is quite comprehensive and covers most of the PSFs listed in NUREG-1792, Good Practices for Implementing Human Reliability Analysis (HRA). PSF considered – Available time, stress/stressors, complexity, experience/training, procedures, ergonomics/HMI, fitness for duty, work processes. A comparison of PSFs across a number of HRA techniques, HEART, CREAM, THERP ASEP is also provided.
		The method does not consider a majority set of factors that affect human reliability.		

Content validity	Essential	**Sub-scale 2** **Quantitative sensitivity**		SPAR-H
		The method is quantitatively sensitive to the effect of each individual PSF considered qualitatively.	X	Justification All of the PSFs have quantitative weights that, if chosen, will have an impact on the HEP.
		The method is not quantitatively sensitive to individual PSFs, but makes a single adjustment to the HEP based on the contribution of the overall combination of the PSFs considered.		
		The method is not quantitatively sensitive to PSFs.		
		Sub-scale 3 **Interaction between factors** Typically HRA methods adopt a linear multiplicative combination of PSFs. It is recognised that some PSFs may interact in other ways, e.g. a step change in the effect of one PSF once a threshold has been reached on a second PSF, or where the effect of the combination of two PSFs is far greater than multiplicative relationship would predict or where one PSF has a triggering effect on other PSFs in a causal chain.		SPAR-H
		Interactions between PSFs are accounted for on the basis of knowledge of the relationship between specific PSFs.		Justification As an adjustment, if more than three PSFs have impact, this is adjusted for to lessen the impact. Exponential effects in which PSFs interact and make the HEP higher than accounted for in the linear model are not included in SPAR-H. The PSFs in SPAR-H, as in any PSF based HRA method that includes a comprehensive set of PSFs, are not completely independent. This has to be evaluated by the assessor in every analysis. They do provide a table (G-1, section 2.7.5, appendix G) in which they qualitatively describe relations between PSFs, so the analyst can use this in their expert judgment.
		Combinations of PSF effects are accounted for using a simple linear model.	X	
		Interactions between or combination of PSF effects are not considered by the method.		

Content validity	Essential	**Attribute 7** **Consideration of human error dependency** Modelling should include consideration of human error dependencies or common cause failures.		SPAR-H
		The method provides a procedure for identifying potential sources of dependence among Human Failure Events (HFEs) and/or sub-tasks of an HFE, and provides a method to derive conditional HEPs based on the systematic assessment of these sources of dependence.	X	Justification SPAR-H uses a clarified and "mechanistic" version of the THERP dependency model, with a list of factors to be evaluated in order to determine the level of dependence that should be applied. Mathematically this is clear.
		The method identifies potential sources of dependence, but does not provide a process for linking these sources of dependence to a quantified model for deriving conditional HEPs.		
		The method does not address dependencies and common cause mechanisms among tasks and sub-tasks.		

Content validity	Essential	**Attribute 8** **Consideration of deviations and progressions in accident sequences** The method should provide a capability to accommodate: • Deviations from nominal accident scenarios due to: (A) Plant conditions: 1. Aleatory factors, such as sizes and locations of equipment failures and time sequences. 2. Complicating factors, such as coincident failures in control, instrumentation and support systems not normally modelled explicitly in PSA models. (B) Human failure scenarios due to crew organisational & operating practices that introduce opportunities to create new failure mechanisms. • Fault progressions including consequential faults and accident sequences encompassing Level 1 and Level 2 PSA which may involve extended time sequences and degraded operating environments should also be accommodated.	SPAR-H

Content validity	Essential	**Sub-scale 1** **Deviations**		SPAR-H
		The method provides for the qualitative and quantitative assessment of all the types of deviations in accident scenarios.		Justification The method can in principle be used to any kind of combination of faults and sequences. This has to be reflected by the analyst though. SPAR-H does not support qualitative analysis. SPAR-H provides for the quantitative assessment of deviations in accident sequences. Modelling conventions are discussed in section 4.1. However, SPAR-H does not provide a method for qualitative decomposition of HEPs. Rather, it suggests that the analyst use the 10 step method derived from the ATHEANA method. Section 2.7.7 describes how non-SPAR-H information can be combined with SPAR-H information in a single HEP. As many recovery paths as needed may be included.
		The method provides for the qualitative assessment of human error during fault progressions, but does not provide for the derivation of HEPs in support of this assessment.		
		The method does not provide a means to deal with deviations in accident scenarios	X	
		Sub-scale 2 **Fault progression**		SPAR-H
		The method provides for the qualitative and quantitative assessment of human errors during fault progressions including level 1 to level 2 PSA fault progressions		Justification The method can in principle be used for any kind of sequences, but it does not include the modelling aspects. Note: the intermediate rating should be described opposite for SPAR-H, because it does not provide for the qualitative modelling, but provides for the derivation of HEPs given a qualitative model. In terms of long time horizons associated with events, SPAR-H covers this in terms of the ratio of time required to time available, one of the performance shaping factors. Opportunities for additional errors of the same or different type due to changes in timing would be characterized in suitable logic structures such as fault trees and then quantified. For the discovery process, the analyst is directed to the ATHEANA process or best practices as a substitute for the ATHEANA search process. SPAR-H provides for quantitative assessment during fault progressions, and can be applied in level 1 & 2 PSA fault progressions. SPAR-H has been applied to LP/SD operations (see Section 5.1).
		The method provides for the qualitative assessment of human error during fault progressions, but does not provide for the derivation of HEPs in support of this assessment.		
		The method does not provide for the qualitative and quantitative assessment of human errors during fault progressions.	X	

Content validity	Highly desirable	**Attribute 9** **Consideration of cognitive error** The method should be sensitive to the factors that influence the diagnosis and decision making component of the response to an initiating event.		SPAR-H
		The method estimates the probability of cognitive error based on the assessment of a set of factors that are known to affect diagnosis and decision making performance	X	Justification Specific work sheets are given for diagnosis that can be used on any type of cognitive activity, also within procedure execution.
		The method uses a simple model such as a time reliability curve as the primary factor for estimating the probability of cognitive error**.**		
		The method provides no way of estimating the likelihood of cognitive error**.**		

Content validity	Highly desirable	**Attribute 10** **Consideration of statistical uncertainty** The method should provide for statistical uncertainty analysis of derived human error probabilities.		SPAR-H
		The method derives uncertainty parameters from experience (either in-plant or from relevant simulator trials).		Justification The user manual (NUREG/CR-6883) discusses uncertainty. Uncertainty values are derived using a Beta distribution rather than the log-normal distribution that is used in techniques such as THERP and NARA.
		The method provides generic uncertainty parameters, e.g. standardised error factors.	X	
		The method provides no uncertainty parameters.		

Content validity	Desirable	**Attribute 11** **Consideration of organisational issues** The method should consider the impact of organisational issues including safety-culture factors (attitudes and behaviours), and organisational process factors (e.g. command and control structures, conflicts of interest, communication and decision making protocols on human reliability).		SPAR-H
		Sub-scale 1 **Safety-culture factors** (attitudes and behaviours).		SPAR-H
		The method provides an adequate quantitative method to adjust HEPs based on an assessment of safety culture/safety climate**.**		Justification The PSFs include work processes and fitness for duty. The way in which these are used for safety culture is up to expert judgment though.
		The method provides a qualitative means to assess safety culture/safety climate, but does not include a process to modify HEPs based on the assessment**.**		
		The method does not take into account safety culture factors**.**	X	
		Sub-scale 2 **Process factors** (e.g. command and control structures, communication and decision making protocols on human reliability).		SPAR-H
		The method provides a quantitative method to assess process factors	X	Justification The PSFs include work processes and fitness for duty. Work processes deals with factors such as shift hand over, communication, command and control, other organisational issues etc. However, work processes is a "catch all" PSF, may be too wide.
		The method provides a qualitative means to assess process factors, but does not include a process to modify HEPs based on the assessment**.**		
		The method does not take into account process factors**.**		

Empirical validity	Essential/Desirable	**Attribute 12** **Empirical validity** The method should demonstrate evidence of empirical validation exercises, peer review processes or community acceptance based on application and maturity.		SPAR-H
		Sub-scale1 **Statistical evidence**		SPAR-H
		The HEP estimates derived by the method have been shown to demonstrate good agreement with plant and /or simulator data for comparable tasks**.**		Justification SPAR-H was part of the HRA empirical study in Halden (NUREG/IA-0216 vol. 1,2,3), and the results were mixed. However due to the non-statistical treatment of the data generated by the international empirical study, it is not considered to provide evidence in relation to this attribute in this study. Section 3, Analysis, discusses how SPAR-H was validated. As a check on the validity of the PSFs, the INL reviewed operating events from Human Performance Event Database (HPED) to identify instances where the effects of SPAR-H PSFs could be identified. HPED contains reviews of LERs and AIT reports. The effects of PSFs were consistent with the available data. The diagnosis base rate used in SPAR-H is within the range used by other methods. In section 3.2, the validation of SPAR-H by an analysis team is discussed. Section 3.4 discusses how the At Power worksheets were used in field testing on NASA processes. Thus, there has been a part validation of the PSFs and convergent validity test of the base values.
		The HEP estimates derived by the method have been shown to demonstrate good agreement with HEP estimates produced by other HRA methods for the same or comparable tasks**.**		
		The method has failed to derive comparable HEP estimates in tests of empirical validity or has not been subject to such assessments**.**	X	

Empirical validity	Essential/Desirable	**Sub-scale 2** **Verification/Peer review**		SPAR-H
		The method has been subject to peer review by a team of recognised HRA experts, and the peer review comments have been incorporated to the development of the method**.**	X	Justification Appendix I, NUREG/CR-6883 describes a thorough peer review.
		The method has been subject to peer review by a single, recognised HRA expert, and the comments have been incorporated to the development of the method**.**		
		The method has not been subject to independent peer review or the method has not been updated in response to peer review comments.		
		Sub-scale 3 **Application/Maturity**		SPAR-H
		The method has been extensively applied, internationally, for five or more years**.**	X	Justification Mainly used by the NRC, but also now included in the EPRI calculator. International use: AREVA UK
		The method has been applied to a limited number of HRAs**.**		
		The method has not yet been applied to a HRA**.**		

Reliability	Essential	**Attribute 13** **Computer models and software tools** If a method incorporates the use of a computer model or software tool to analyse a human action, A QA programme should be applied to ensure quality of the design and validity of the output.		SPAR-H
		A relevant, recognised/accepted international standard has been applied to the software design and verification of the computer based HRA method/tool**.**		Justification N/A. Paper based sheets.
		The design of the computer based HRA method/tool is based upon a documented QA process, which includes software verification**.**		
		There is no evidence that the design of the computer based HRA method/tool is based on a structured and validated software development or QA method that includes software verification**.**		
Reliability	Highly desirable	**Attribute 14** **Reliability and traceability** The method should provide consistent qualitative and quantitative information for comparable scenarios within analysts and between analysts for similar scenarios. The method should also provide sufficient information to facilitate tracing estimates back to input assumptions.		SPAR-H
		Sub-scale 1 **Within analyst consistency/reliability**		SPAR-H
		A formal comparison, amenable to statistical analysis, has been undertaken to demonstrate that the same HRA analyst provides consistent answers for analyses made at different times for the same scenario**.**		Justification There is no such evaluation in the public domain.
		An informal comparison has been undertaken, which suggests good within analyst agreement for analyses made at different times**.**		
		There is no information available to suggest good within analyst agreement for analyses made at different times**.**	X	

		Sub-scale 2 **Between analyst consistency/reliability**		SPAR-H
Reliability	Highly desirable	A formal comparison, amenable to statistical analysis, has been undertaken to demonstrate that different HRA analysts provide consistent answers for the same scenario**.**		Justification In the SPAR-H development process, a limited formal comparison among analysts was conducted, yielding positive findings. However, this is not available in the public domain.SPAR-H. More recently, SPAR-H was assessed in both the International [Reference 8 of this NEA document, page 57] (NUREG-2127) and U.S. HRA Empirical Studies [Reference 9 of this NEA document, page 57]. The U.S. HRA Empirical Study addressed the consistency/reliability of analyses performed by different analysis teams with the ASEP method (as well as other analysis teams applying CBDT+HCR/ORE, SPAR-H, and ATHEANA). Although the results with ASEP (as well as with the other methods) found some consistency in the HEPs obtained for the HFEs examined in the study, a detailed comparison of the HRAs of the HFEs found significant differences in the qualitative findings used by the analysis teams to estimate the HEPs. Furthermore, there were also significant differences in the assessed contribution of the diagnosis/decision and execution components of the HEPs. Differences in the implementation of SPAR-H by the analysis teams, e.g. one team treating the HFE holistically while the other team decomposed the HFE into subtasks that were individually quantified by applying SPAR-H, also contributed to these differences in qualitative and quantitative predictions. Consequently, in spite of the limitations of the U.S. HRA Empirical Study, its results suggest that the between-analyst consistency is superficial. In conclusion, this information does not suggest good between-analyst agreement.
		An informal comparison has been undertaken, which suggests good between analyst agreement**.**		
		There is no information available to suggest good between analyst agreement**.**	X	

Reliability	Highly desirable	**Sub-scale 3** **Traceability**		SPAR-H
		The method provides a procedure to ensure easy, complete traceability of the estimates of human performance in the HRA, such that an independent reviewer could trace back HEPs to relevant assumptions, models and data cited in the method**.**	X	Justification Traceability has two sides in PSF based methods: 1. Traceability of the quantification itself, how the HEP is constructed given the choice of the PSF weights. This is very good and easy in SPAR-H 2. Traceability of the assumptions on which the choices of the PSF weights are made. In SPAR-H, this depends on good documentation of the qualitative analysis, and is dependent on the analyst. The method does not force analysts to be good on this one. The high rating is assigned because item 1 is valid for this point. 2) is a function of the analysis.
		The HRA method itself does not provide a procedure for traceability, but there is sufficient information available about the method to facilitate traceability, and enable an independent reviewer to understand what was done**.**		
		There is insufficient information available to facilitate traceability**.**		

Usability	Highly desirable	**Attribute 15** **Definition of method scope** The scope of the method should be clearly defined.		SPAR-H
		The scope of the method is clearly defined in a user manual and/or technical basis document**.**	X	Justification NUREG/CR-6883 identifies that SPAR-H can be used for Type A and C HFEs and can be used either for screening or best estimate applications. (Note: It is questionable whether SPAR-H is good for screening though, since it has no mechanism to ensure conservative HEPs.)
		The scope of the method is described vaguely and some analyst judgement is required to determine its applicability to a particular human action/error**.**		
		The scope of the method is not defined**.**		

Usability	Highly desirable	**Attribute 16** **Qualitative outputs** The method should produce qualitative outputs that are useful to inform human factors and safety management improvements at the plant		SPAR-H
		The method generates qualitative information to inform improvements to reduce the potential for human error that is explicitly related to each of the factors that are used in the method to derive an HEP.		Justification This very much depends on the documentation of the qualitative analysis and its input to the PSF weighting. It is not enough to score the PSF in order to give input to the error reduction. This may be on a too high level than what is useful for the plants.
		The method generates qualitative information to inform improvements to reduce the potential for human error, but this is not explicitly linked to each of the factors used in the derivation of HEPs.	X	Also, one may not find the potential error reduction at all if the PSFs are analysed on a too high level in the scenario. One must dive into the details of the scenario to get to the potentials for errors, and SPAR-H does not force the analyst to do this.
		The method does not generate qualitative information to inform improvements to reduce the potential for human error.		

Usability	Highly desirable	**Attribute 17** **Qualitative uncertainty and quantitative conservatism** Methods should be able to reflect uncertainties related to qualitative information via conservatisms in the quantification process.		SPAR-H
		The method provides a mathematical procedure for adjusting the conservatism of the HEPs derived as a function of the level of certainty in the qualitative information collected during the assessment.		Justification If the PSFs are scored based on lacking qualitative information, this is not necessarily noted or has any impact on the HEPs, especially not in a pessimistic direction. Qualitative uncertainty may result in optimistic values. The reason for this is that if there is insufficient information, a nominal value is assumed on the PSFs.
		The method provides a general caution on the need to adjust the conservatism of HEPs as a function of the level of certainty in the qualitative information collected, but does not provide a mathematical procedure for doing so.		This is also the reason that SPAR-H should not be used for screening.
		The method does not address the issue of uncertainties in qualitative information and the impact of this on derived HEPs.	X	

Usability	Desirable	**Attribute 18** **Availability of user documentation** The method should be supported by a detailed user documentation e.g., manual or instructions, which describes how the method should be applied.		SPAR-H
		The method contains user documentation that provides a detailed step-by-step procedure for all steps in the derivation of an HEP.	X	Justification NUREG/CR-6883 is comprehensive. It could be improved on some detailed guidance on how to weight PSFs and on requirements to the degree of details required in the qualitative analysis. There is also a recent companion document INL/EXT-10-18533 Rev 2 SPAR-H step-by-step guidance.
		The method contains user documentation that provides a high level description of how it is applied to derive HEPs, but not all elements of the method are detailed as step-by-step procedures.		
		The method provides only a high level description of its method of application and or data tables for the derivation of HEPs.		

Usability	Desirable	**Attribute 19** **Use of limiting values** The method should provide limiting values. (Relevant Good Practice documents discuss limiting values that are used in member countries).		SPAR-H
		The method provides limiting values and advice on their application.	X	Justification In the user manual, it is stated that: "A lower bound cut-off of 1.0E-5 for HEPs is suggested." However, this is not repeated in the worksheets, so it may easily be overlooked. It is mentioned in the extension user's guide.
		The method provides advice on the need to limit claims on human performance but does not provide specific limiting values.		
		The method does not consider the use of limiting values.		

Usability	Indifferent/Essential	**Attribute 20** **Resources** A comparative estimate of the resources (time, cost, utility demands, level of specialist training required, level and type of knowledge required) needed to apply the method in comparison with other HRA methods.		SPAR-H
		The estimated cost of and time required for applying the HRA method is less than or comparable with that of other HRA methods**.**	X	Justification SPAR-H can be used with very little resources. This is also a danger, if it used in a very simple way, the results may lack confidence.
		The estimated cost of and time required for applying the HRA method is in excess of that required for application of other HRA methods**.**		SPAR-H was developed to be a low cost, easy to apply method. It assumes that the analyst has significant operations experience, but little experience with human performance. The estimated cost and time to utilize SPAR-H is less than most other HRA methods. The documentation in NUREG/CR-6883 and some basic background in PSA or HRA is thought to be sufficient for employing the method. Note: The topic of a 2.5 day training course for SPAR-H has been discussed. To date SPAR-H has been taught as part of a survey of HRA methods presented to NRC staff.

A2.7 Attribute Evaluations – HCR/ORE & CBDT

Desirable Attributes of HRA – Methods Evaluation Scale – HCR/ORE & CBDT

Instructions to assessors

Indicate your evaluation of how well a method meets the requirements of an attribute by placing a tick in the relevant coloured box.

Provide a succinct justification for the rating you have allocated in the text box labelled justification.

You should complete an evaluation for each attribute and each sub-scale of an attribute.

If an attribute is not relevant to the method you are evaluating then record this in the justification box with an explanation as to why the attribute is not relevant.

Your evaluation should be based on the application of the method exactly as it is described in the method's user documentation. If there are modifications to the process which improve its application, but which are not formally recorded in an update to the method's documentation, then these can be noted in the justification box as potential improvements, but they should not be considered as part of the evaluation.

Within the method evaluation scale, a high rating (dark blue) indicates that the requirements of an attribute have been fully or largely met for the method's intended scope of application. An intermediate rating (medium blue) indicates that a method meets some, but not all of the requirements of the attribute. A low rating (light blue) indicates that the requirements of the attribute are not met or that no evidence is available in relation to the attribute for the method.

Note: Where HCR/ORE and CBDT have different ratings:
- "H" is used to denote the HCR/ORE rating and
- "C" the CBDT rating.

Construct validity	Essential	**Attribute 1** **Availability of information relating to the technical basis of the method** Information is provided on the technical basis of the method, in terms of its scientific underpinnings and data, in order to allow a judgement on the validity of the method to be made.		HCR/ORE & CBDT
		Comprehensive information on the technical basis and/or data underpinning the method is available and its application is discussed as part of the documentation of the method.	X	Justification The review is based on EPRI TR-100259" An Approach to the Analysis of Operator Actions in Probabilistic Risk Assessment" (June 1992), EPRI NP-6937 "Operator Reliability Experiments Using Power Plants Simulators" (Vol. 1, 2, 3, July 1990 and January 1991) and EPRI TR-101711 "SHARP 1- A Revised Systematic Human Action Reliability Procedure" (T 1 and T 2, December 1992). Also relevant is "Establishing minimum acceptable values for probabilities of human failure events – practical guidance for Probabilistic Risk Assessment", interim report, Report 1021081, October 2010 Access to the EPRI HRA Calculator software tool was not available.
		The method provides references that allow the information forming the technical basis and/or the data underpinning it to be obtained.		
		The method does not provide sufficient information to allow its technical basis and underpinning data to be accessed for review.		

Construct validity	Essential	**Attribute 2** **The technical basis of the method (Theory)** The technical basis of the method is based upon, and does not contradict, a relevant body of scientific knowledge		HCR/ORE & CBDT
		The method operationalises a relevant model of human performance or system safety which has scientific acceptance.	C	Justification As described in EPRI TR-100259, the focus of the approach is the HCR/ORE quantification method, with recognition of the CBDT supplementary method. In addition the THERP model is used for the execution component of HFEs. A critical element of the HCR/ORE method is the use of simulator data. The authors claim that as long as the simulator is a faithful representation of the plant response, many of the important performing shaping factors are addressed. The HCR/ORE methodology assumes that the error probability is a function of the "normalised" time (ratio between allowable time window and observed average operator response time). As a Time-Reliability Curve, the method has a certain face validity. On the other hand, the "normalisation" of the time window distinguishes this method from other Time-Reliability Curves. The theoretical basis for this basic assumption is limited. In practice, large ratios lead to very low human error probabilities that may not be credible when the allowable time window is small. This problem is claimed to be solved through the use of the supplementary CBDT method for failure probabilities below 0.01 based on a small number of simulator data.
		Elements of the method are inconsistent with an accepted scientific model of human performance or system safety.	H	

Construct validity	Essential	**Attribute 3** **The technical basis of the method (Data)** Where the technical basis of the method is based on a dataset, the source of the data/information and its relevance for application in the nuclear industry should be demonstrated.		HCR/ORE & CBDT
		The data underlying the method are largely based on observations of actual or simulated task performance in nuclear industry tasks.	H	Justification For the HCR/ORE method, the time reliability correlation is based on plant-specific simulator data (observations of crew performances) . However, it should be noted that the part of the correlation from which the failure probability is derived is largely based on an extrapolation of the observed success data. In principle, the HCR/ORE method can and is best applied with data from the plant being analysed. In practice plant HRA analyses are almost always based on the generic data. Consequently, a) some modifications to the correlation should be necessary to transfer them to a different environment (e. g. differences in control room lay-out, staffing, operational practice, and so on); and b) the data reflect potential "simulator biases" (such as those listed in NEA/CSNI R(98)1 (Critical Operator Actions – Human Reliability Modelling and Data Issues, February 1998). For the CBDT method, an "orange" is assigned because most of the probabilities assigned to the endpoints are based on judgement and derived from THERP as documented in EPRI NP-69373. The THERP data is itself derived from a combination of nuclear and non-nuclear data and expert judgement.
		The data underlying the method are based on expert judgement or observations of human performance for relevant tasks in a domain that is closely related to the nuclear industry e.g. other high hazard industries.	C	
		The data underlying the method are taken from tasks that are not related or relevant to nuclear industry tasks.		

Construct validity	Highly desirable	**Attribute 4** **Internal consistency of the method** The method demonstrates internal consistency between the technical basis, the error definition, the PSFs and the qualitative and quantitative method steps		HCR/ORE & CBDT
		The qualitative and quantitative component parts of the method are theoretically compatible and form a coherent consistent whole.	X	Justification HCR/ORE and CBDT are intended to be used in combination; the internal consistency of the method should consider the internal consistency of each component and that of the overall method. The individual components, HCR/ORE and CBDT, are each internally coherent. Considering the components jointly, they are designed to be complementary so that the overall method is coherent.
		There are theoretical inconsistencies between the qualitative and/or quantitative components of the method.		

Content Validity	Highly desirable	**Attribute 5** **Qualitative assessment** It is recognised good practice that HRA quantification is supported by qualitative analysis to develop an understanding of operator performance within the scenario that is being assessed. This attribute considers the extent to which the qualitative analysis stages of the HRA (e.g. task analysis and error identification) is directed or prescribed by the HRA method, beyond providing a set of performance shaping factors to be considered.	HCR/ORE & CBDT	
		The method contains or prescribes a process for conducting qualitative assessment**.**		Justification EPRI TR-100259 contains no explicit, formal guidance nor references any such guidance for qualitative analysis. The EPRI SHARP1 framework is mentioned but this guidance is for a high-level process rather than detailed guidance on performing qualitative analysis.
		The method includes a general statement indicating that a qualitative assessment should be provided, e.g. by referring to the use of task analysis**.**	X	
		The method does not make any reference to qualitative analysis**.**		

Content validity	Essential	**Attribute 6** **Factors influencing human reliability considered by the method** The method should be quantitatively sensitive to a majority of accepted factors* (PSFs) that influence human reliability. *: There are pre-defined lists of PSFs/EPCs available throughout the literature and within HRA methods (typically 1st generation methods). This attribute does not seek to define such a list, so as to accommodate developments in human performance, system safety and accident analysis. The evaluation teams are asked to use professional judgment when considering this attribute. It is not expected that methods will be assigned a high rating (dark blue) if only a small number of factors are accommodated, but we do not prescribe an absolute number of factors that are required.	HCR/ORE & CBDT	
		Sub-scale 1 **Adequacy of PSFs**.		
		The method requires qualitative assessment of a majority of accepted factors that affect human reliability (PSFs).	X	Justification For this attribute, the adequacy of the set of factors considered by the method as a whole (HCR/ORE + CBDT) is assessed. It is noted however that nearly all of the factors are treated explicitly in the CBDT component of the method. The set of factors used in the CBDT method can be mapped to the set of performance shaping factors (as required by the ANS/ASME PRA Standard (RA-Sa-2009) in addition to all PSFs outlined in NUREG-1921, NUREG-1792, and NUREG-1842.
		The method does not consider a majority set of factors that affect human reliability.		

Content validity	Essential	**Sub-scale 2** **Quantitative sensitivity**		HCR/ORE & CBDT
		The method is quantitatively sensitive to the effect of each individual PSF considered qualitatively.	C	Justification The CBDT method produces different quantitative results for different assessments of those PSFs considered in the model (i.e., assessed "high"). In the application of HCR/ORE, the effect of the PSFs on a task modeled by an HFE is addressed by the sigma factor. The method explicitly treats the PSF related to the adequacy of time, through the normalized time ratio. In practice, if HCR/ORE is applied in the usual way by estimating sigma and median time to response, there is a lack of guidance for considering the effect of the PSFs on sigma and the median time to response.
		The method is not quantitatively sensitive to individual PSFs, but makes a single adjustment to the HEP based on the contribution of the overall combination of the PSFs considered.		
		The method is not quantitatively sensitive to PSFs.	H	
		Sub-scale 3 **Interaction between factors** Typically HRA methods adopt a linear multiplicative combination of PSFs. It is recognised that some PSFs may interact in other ways, e.g. a step change in the effect of one PSF once a threshold has been reached on a second PSF, or where the effect of the combination of two PSFs is far greater than multiplicative relationship would predict or where one PSF has a triggering effect on other PSFs in a causal chain.		HCR/ORE & CBDT
		Interactions between PSFs are accounted for on the basis of knowledge of the relationship between specific PSFs.		Justification As noted above, the HCR/ORE method is insensitive to PSFs and is assigned a low rating. In the CBDT method, there are two failure modes (failures of the plant information-operator interface, and failures in the procedure-crew interface). For each failure mode, four failure mechanisms (different for each mode) are used to identify "causes" that are related to or influenced by PSF-like qualitative assessments. By using a decision tree for each mechanism, therefore, some interaction between PSFs is modelled for the failure mechanism represented. This is considered equivalent to an orange rating. However, it should be noted that, in practice, such overlaps can lead to difficulties in users deciding which decision tree to use to represent a specific qualitative analysis output.
		Combinations of PSF effects are accounted for using a simple linear model.	C	
		Interactions between or combination of PSF effects are not considered by the method.	H	

Content validity	Essential	**Attribute 7** **Consideration of human error dependency** Modelling should include consideration of human error dependencies or common cause failures.		HCR/ORE & CBDT
		The method provides a procedure for identifying potential sources of dependence among Human Failure Events (HFEs) and/or sub-tasks of an HFE, and provides a method to derive conditional HEPs based on the systematic assessment of these sources of dependence.		Justification EPRI TR-100259 includes some discussion of dependency (including timing issues that can be sources of dependency) in Section 7.2 and refers to SHARP1 (i.e., assigned "orange"). SHARP 1 provides a dependency assessment model. The method recommends the application of the THERP dependency model to assess potential dependencies between HFEs.
		The method identifies potential sources of dependence, but does not provide a process for linking these sources of dependence to a quantified model for deriving conditional HEPs.	X	
		The method does not address dependencies and common cause mechanisms among tasks and sub-tasks.		

Content validity	**Essential**	**Attribute 8** **Consideration of deviations and progressions in accident sequences** The method should provide a capability to accommodate: • Deviations from nominal accident scenarios due to: (A) Plant conditions: 1. Aleatory factors, such as sizes and locations of equipment failures and time sequences. 2. Complicating factors, such as coincident failures in control, instrumentation and support systems not normally modelled explicitly in PSA models. (B) Human failure scenarios due to crew organisational & operating practices that introduce opportunities to create new failure mechanisms. • Fault progressions including consequential faults and accident sequences encompassing Level 1 and Level 2 PSA which may involve extended time sequences and degraded operating environments should also be accommodated.	HCR/ORE & CBDT

Content validity	Essential	**Sub-scale 1** **Deviations**		HCR/ORE & CBDT
		The method provides for the qualitative and quantitative assessment of all the types of deviations in accident scenarios.		Justification Because EPRI TR-100259 does not provide an explicit, formal qualitative analysis approach, it has no way to identify deviations in accident scenarios. If a separate qualitative analysis were done, the CBDT approach may be able to represent certain aspects of such deviation scenarios. The HCR/ORE method is based on a limited amount of simulator observations including a limited amount of factors influencing human behaviour and performance. Therefore additional experiments would be necessary to quantify deviations assuming that the user is able to identify and evaluate deviations and to run new experiments.
		The method provides for the qualitative assessment of human error during fault progressions, but does not provide for the derivation of HEPs in support of this assessment.		
		The method does not provide a means to deal with deviations in accident scenarios	X	
		Sub-scale 2 **Fault progression**.		HCR/ORE & CBDT
		The method provides for the qualitative and quantitative assessment of human errors during fault progressions including level 1 to level 2 PSA fault progressions		Justification The EPRI developers state that “the method in the EPRI HRA approach support a Level 1 PSA”. An extension of the method to support a Level 2 PSA is reportedly under development at this time.
		The method provides for the qualitative assessment of human error during fault progressions, but does not provide for the derivation of HEPs in support of this assessment.		
		The method does not provide for the qualitative and quantitative assessment of human errors during fault progressions.	X	

Content validity	Highly desirable	**Attribute 9** **Consideration of cognitive error** The method should be sensitive to the factors that influence the diagnosis and decision making component of the response to an initiating event.		HCR/ORE & CBDT
		The method estimates the probability of cognitive error based on the assessment of a set of factors that are known to affect diagnosis and decision making performance**.**	C	Justification The HCR/ORE method considers cognitive error via the use of a TRC Through the use of decision trees of the CBDT approach, a limited number of factors related to cognitive failure are addressed.
		The method uses a simple model such as a time reliability curve as the primary factor for estimating the probability of cognitive error**.**	H	In common with other 1st generation HRA methods, the method does not address a number of additional important determinants of the cognitive error that are addressed in recent methods that focus on cognitive error in more detail.
		The method provides no way of estimating the likelihood of cognitive error**.**		

Content validity	Highly desirable	**Attribute 10** **Consideration of statistical uncertainty** The method should provide for statistical uncertainty analysis of derived human error probabilities.		HCR/ORE & CBDT
		The method derives uncertainty parameters from experience (either in-plant or from relevant simulator trials).		Justification EPRI TR-100259 provides minimal discussion on uncertainty and generally refers to other work.
		The method provides generic uncertainty parameters, e.g. standardised error factors.	X	Standardised error factors are applied to the final human error probability in the EPRI HRA approach.
		The method provides no uncertainty parameters.		The HCR/ORE methodology provides explicitly uncertainty parameters (Table 3-1 of the reviewed document and also comment on attribute 17). The failure probabilities assigned to the CBDT are based on THERP, including the generic uncertainties.

Content validity	Desirable	**Attribute 11** **Consideration of organisational issues** The method should consider the impact of organisational issues including safety-culture factors (attitudes and behaviours), and organisational process factors (e.g. command and control structures, conflicts of interest, communication and decision making protocols on human reliability).		HCR/ORE & CBDT
		Sub-scale 1 **Safety-culture factors** (attitudes and behaviours).		HCR/ORE & CBDT
		The method provides an adequate quantitative method to adjust HEPs based on an assessment of safety culture/safety climate**.**		Justification The method does not explicitly take into account safety culture factors. The EPRI CBDT does include a failure mechanism for "deliberate violations" to capture instances where the operators do not believe the procedures fit the situation.
		The method provides a qualitative means to assess safety culture/safety climate, but does not include a process to modify HEPs based on the assessment**.**		
		The method does not take into account safety culture factors**.**	X	
		Sub-scale 2 **Process factors** (e.g. command and control structures, communication and decision making protocols on human reliability).		HCR/ORE & CBDT
		The method provides a quantitative method to assess process factors		Justification Some process factors are implicitly embedded in the data gained by the simulator experiments but the data cannot be traced back to those factors. In general, the method does not take into account work process factors although it does allow for the adjustment of time (the estimated median response time) to account for command and control structures (or effects such as delays in communications.
		The method provides a qualitative means to assess process factors, but does not include a process to modify HEPs based on the assessment**.**		
		The method does not take into account process factors**.**	X	

Empirical validity	Essential/Desirable	**Attribute 12** **Empirical validity** The method should demonstrate evidence of empirical validation exercises, peer review processes or community acceptance based on application and maturity.		HCR/ORE & CBDT
		Sub-scale1 **Statistical evidence**		HCR/ORE & CBDT
		The HEP estimates derived by the method have been shown to demonstrate good agreement with plant and /or simulator data for comparable tasks**.**		Justification The HCR/ORE & CBDT was evaluated in the International and US HRA Empirical Studies (NUREG/IA-0216 vol. 1,2,3); in these studies, the HEPs estimated by the analysis teams were assessed against a set of HEP confidence bounds for the HFEs of interest, which were derived from the observed performances of crews on simulators. While the results were moderately positive, it should be noted that the Empirical Studies are not quantitative validation studies.
		The HEP estimates derived by the method have been shown to demonstrate good agreement with HEP estimates produced by other HRA methods for the same or comparable tasks**.**		
		The method has failed to derive comparable HEP estimates in tests of empirical validity or has not been subject to such assessments**.**	X	
		Sub-scale 2 **Verification/Peer review**		HCR/ORE & CBDT
		The method has been subject to peer review by a team of recognised HRA experts, and the peer review comments have been incorporated to the development of the method**.**		Justification Documentation of a formal, independent peer re view is not known to the reviewers. A high-level evaluation of the EPRI approach (and other methods) is included in NUREG-1842.
		The method has been subject to peer review by a single, recognised HRA expert, and the comments have been incorporated to the development of the method**.**		
		The method has not been subject to independent peer review or the method has not been updated in response to peer review comments.	X	

Empirical validity	Essential/Desirable	**Sub-scale 3** **Application/Maturity**		HCR/ORE & CBDT
		The method has been extensively applied, internationally, for five or more years**.**	X	Justification EPRI TR-100259 was published in 1992. The EPRI HRA Users Group was started in 2000, and since 2000 the EPRI HRA approach has been used extensively through implementation in the EPRI HRA Calculator for both U.S and international HRA applications.
		The method has been applied to a limited number of HRAs**.**		
		The method has not yet been applied to a HRA**.**		

Reliability	Essential	**Attribute 13** **Computer models and software tools** If a method incorporates the use of a computer model or software tool to analyse a human action, A QA programme should be applied to ensure quality of the design and validity of the output.		HCR/ORE & CBDT
		A relevant, recognised/accepted international standard has been applied to the software design and verification of the computer based HRA method/tool**.**		Justification The EPRI HRA Approach has been captured in a software tool: the EPRI HRA Calculator™. The EPRI HRA Calculator has been developed under a quality assurance (QA) program and continues to be updates and modified. The QA process is based on selected elements of 10CFR50 Appendix B for safety-related software, including a design document and verification testing. The software developer is qualified to develop safety-related software and is audited bi-annually. Every major release of the EPRI HRA Calculator™ is developed following a software specification and then has been subjected to validation and verification testing.
		The design of the computer based HRA method/tool is based upon a documented QA process, which includes software verification**.**	X	
		There is no evidence that the design of the computer based HRA method/tool is based on a structured and validated software development or QA method that includes software verification**.**		

Reliability	Highly desirable	**Attribute 14** **Reliability and traceability** The method should provide consistent qualitative and quantitative information for comparable scenarios within analysts and between analysts for similar scenarios. The method should also provide sufficient information to facilitate tracing estimates back to input assumptions.		HCR/ORE & CBDT
		Sub-scale 1 **Within analyst consistency/reliability**		HCR/ORE & CBDT
		A formal comparison, amenable to statistical analysis, has been undertaken to demonstrate that the same HRA analyst provides consistent answers for analyses made at different times for the same scenario**.**		Justification There have been no formal studies conducted to demonstrate that the same analyst would provide consistent answer for analyses made at different times. However, analyses conducted for similar plants developed at different times have shown good agreement.
		An informal comparison has been undertaken, which suggests good within analyst agreement for analyses made at different times**.**	X	
		There is no information available to suggest good within analyst agreement for analyses made at different times**.**		

Reliability	Highly desirable	**Sub-scale 2** **Between analyst consistency/reliability**		HCR/ORE & CBDT
		A formal comparison, amenable to statistical analysis, has been undertaken to demonstrate that different HRA analysts provide consistent answers for the same scenario**.**		Justification The U.S. HRA Empirical Study [Ref. 9] addressed the consistency/reliability of analyses performed by different analysis teams with the EPRI HRA method, consisting of the combination of CBDT and HCR/ORE (as well as other analysis teams applying CBDT+HCR/ORE, SPAR-H, and ATHEANA). Although the results for this method (as well as with the other methods) found some consistency in the HEPs obtained for the HFEs examined in the study, a detailed comparison of the HRAs of the HFEs found significant differences in the qualitative findings used by the analysis teams to estimate the HEPs. Furthermore, there were also significant differences in the assessed contribution of the diagnosis/decision and execution components of the HEPs. Consequently, in spite of the limitations of the U.S. HRA Empirical Study, its results suggest that the between-analyst consistency is superficial (i.e. the quantitative results (HEPs) are moderately consistent but the underlying rationales differ). In conclusion, this information does not suggest good between-analyst agreement.
		An informal comparison has been undertaken, which suggests good between analyst agreement**.**		
		There is no information available to suggest good between analyst agreement**.**	X	
		Sub-scale 3 **Traceability**		HCR/ORE & CBDT
		The method provides a procedure to ensure easy, complete traceability of the estimates of human performance in the HRA, such that an independent reviewer could trace back HEPs to relevant assumptions, models and data cited in the method**.**	X	Justification Detailed procedures and documentation sheets are part of the reviewed document. For CBDT, it is easy to identify what decision trees and decision tree branches were used if the appropriate information is provided. For HCR/ORE, the only input to the TRC is time available. An independent reviewer could reproduce the results, so it is judged traceable, too.
		The HRA method itself does not provide a procedure for traceability, but there is sufficient information available about the method to facilitate traceability, and enable an independent reviewer to understand what was done**.**		
		There is insufficient information available to facilitate traceability**.**		

Usability	Highly desirable	**Attribute 15** **Definition of method scope** The scope of the method should be clearly defined.		HCR/ORE & CBDT
		The scope of the method is clearly defined in a user manual and/or technical basis document**.**	X	Justification The scope of the EPRI methods is clearly defined in EPRI TR-100259. However, like any software tool, the methods can be misapplied by analysts who do not "check the book". According to EPRI TR-100259, the methodology cannot evaluate pre-initiator activities, activities causing an initiator in case of error and post-initiator activities not described in procedures. To evaluate execution errors (in- and outside central control room), THERP is applied.
		The scope of the method is described vaguely and some analyst judgement is required to determine its applicability to a particular human action/error**.**		
		The scope of the method is not defined**.**		

Usability	Highly desirable	**Attribute 16** **Qualitative outputs** The method should produce qualitative outputs that are useful to inform human factors and safety management improvements at the plant		HCR/ORE & CBDT
		The method generates qualitative information to inform improvements to reduce the potential for human error that is explicitly related to each of the factors that are used in the method to derive an HEP.		Justification The CBDT method provides qualitative information in the form of failure mechanisms and related PSFs used in the assessment. While there is currently no guidance on how to reduce the potential for human errors, the visibility/traceability of the analysis makes the drivers of the HEP readily apparent. This qualitative information concerning the factors and those factors that are drivers can inform improvements. The CBDT method is the part of the method that does produce this information (while the HCR/ORE plays a minor role); consequently, such information is only available if the CBDT analysis is included (and not omitted on the basis of the HCR/ORE contribution).
		The method generates qualitative information to inform improvements to reduce the potential for human error, but this is not explicitly linked to each of the factors used in the derivation of HEPs.	X	
		The method does not generate qualitative information to inform improvements to reduce the potential for human error.		

Usability	Highly desirable	**Attribute 17** **Qualitative uncertainty and quantitative conservatism** Methods should be able to reflect uncertainties related to qualitative information via conservatisms in the quantification process.		HCR/ORE & CBDT
		The method provides a mathematical procedure for adjusting the conservatism of the HEPs derived as a function of the level of certainty in the qualitative information collected during the assessment.		Justification The HCR/ORE & CBDT methods do not address the issue of uncertainties in qualitative information.
		The method provides a general caution on the need to adjust the conservatism of HEPs as a function of the level of certainty in the qualitative information collected, but does not provide a mathematical procedure for doing so.		
		The method does not address the issue of uncertainties in qualitative information and the impact of this on derived HEPs.	X	

Usability	Desirable	**Attribute 18** **Availability of user documentation** The method should be supported by a detailed user documentation e.g., manual or instructions, which describes how the method should be applied.		HCR/ORE & CBDT
		The method contains user documentation that provides a detailed step-by-step procedure for all steps in the derivation of an HEP.	X	Justification A step-by-step procedure is part of the reviewed document. Further improvements of the methodology are reportedly incorporated into the software tool, which has to be purchased and for which the documentation and training materials were not available for this review.
		The method contains user documentation that provides a high level description of how it is applied to derive HEPs, but not all elements of the method are detailed as step-by-step procedures.		
		The method provides only a high level description of its method of application and or data tables for the derivation of HEPs.		

Usability	Desirable	**Attribute 19** **Use of limiting values** The method should provide limiting values. (Relevant Good Practice documents discuss limiting values that are used in member countries).		HCR/ORE & CBDT
		The method provides limiting values and advice on their application.		Justification With respect to the use of the HCR/ORE method, EPRI TR-100259 discusses the lower limit of its appropriate use (i.e., assigned "orange"). The supplementary approach described in chapter 4 (CBDT) should be applied if HCR/ORE results in probabilities < 0.01. While there is no similar discussion for CBDT, the smallest HEP that can be produced from the decision trees in EPRI TR-100259 appears to be about 1E-4. It should be noted that, in 2010, EPRI published a general guidance document, "Establishing minimum acceptable values for probabilities of human failure events – practical guidance for Probabilistic Risk Assessment", interim report, Report 1021081, October 2010.
		The method provides advice on the need to limit claims on human performance but does not provide specific limiting values.	X	
		The method does not consider the use of limiting values.		

Usability	Indifferent/Essential	**Attribute 20** **Resources** A comparative estimate of the resources (time, cost, utility demands, level of specialist training required, level and type of knowledge required) needed to apply the method in comparison with other HRA methods.		HCR/ORE & CBDT
		The estimated cost of and time required for applying the HRA method is less than or comparable with that of other HRA methods**.**	X	Justification The EPRI methodology developers strived to develop a method that is cost effective; this is particularly true if the generic HCR/ORE data are used. The necessary resources increase if the HCR/ORE correlation cannot be used (CBDT model, evaluating execution errors, evaluating task which are out of scope of the methodology). Also, use of the EPRI HRA Calculator software tool can be helpful in documenting the analysis. However, no comparative cost studies have been performed to make a claim that the EPRI method is cost effective compared to other HRA methods. In EPRI TR-100259 it is highly recommended that analysts produce plant-specific Sigma values for HCR/ORE and CBDTM values by collecting simulator data. Significant resources would be required if the user would like to establish a plant specific qualitative database. However, the report does provide generic data and is widely used by HRA analysts. In order to generate an HEP value the HRA analyst can used the generic values without detailed simulator observations. The HCR/ORE approach is an engineering-based approach. A detailed understanding of human factors is generally not required to apply the method. However, a more robust analysis is often produced when a human factors specialist is included on the team. The rating represents the general use but there can be analyses where the low rating would apply. Specifically, the user has to decide whether or not the HCR/ORE parameters can be applied. Therefore the boundary conditions of HCR/ORE database have to be compared with those impacting the tasks which have to be evaluated. The use of CBDT or the methodologies recommended to assess execution error require specialist Human Factors knowledge While analysts can apply the method without any training, it is ***very highly recommended*** that analysts take one or more EPRI sponsored training course and be supplemented by team members that know the plant response (procedures and success criteria) and know the PRA. This is especially important since there appear to be some important differences between the report guidance and how the approach is applied via the EPRI HRA Calculator and its associated training. The application of any HRA methodology requires a familiarisation and training phase.
		The estimated cost of and time required for applying the HRA method is in excess of that required for application of other HRA methods**.**		

A2.8 Attribute Evaluations – CREAM

Desirable Attributes of HRA – Methods Evaluation Scale – CREAM

Instructions to assessors

Indicate your evaluation of how well a method meets the requirements of an attribute by placing a tick in the relevant coloured box.

Provide a succinct justification for the rating you have allocated in the text box labelled justification.

You should complete an evaluation for each attribute and each sub-scale of an attribute.

If an attribute is not relevant to the method you are evaluating then record this in the justification box with an explanation as to why the attribute is not relevant.

Your evaluation should be based on the application of the method exactly as it is described in the method's user documentation. If there are modifications to the process which improve its application, but which are not formally recorded in an update to the method's documentation, then these can be noted in the justification box as potential improvements, but they should not be considered as part of the evaluation.

Within the method evaluation scale, a high rating (dark blue) indicates that the requirements of an attribute have been fully or largely met for the method's intended scope of application. An intermediate rating (medium blue) indicates that a method meets some, but not all of the requirements of the attribute. A low rating (light blue) indicates that the requirements of the attribute are not met or that no evidence is available in relation to the attribute for the method.

Construct validity	Essential	**Attribute 1** **Availability of information relating to the technical basis of the method** Information is provided on the technical basis of the method, in terms of its scientific underpinnings and data, in order to allow a judgement on the validity of the method to be made.		CREAM
		Comprehensive information on the technical basis and/or data underpinning the method is available and its application is discussed as part of the documentation of the method.		Justification The technical basis of the method in large part is described in detail. Description of data selection and their assignment to cognitive function failures (CFFs) is not sufficient, this accounts for the assignment of the intermediate rating. There is only one paragraph devoted to the issue of data sources in the CREAM guidebook* -- see the discussion related to Attribute 3. Hollnagel, E. (1998). *Cognitive Reliability and Error Analysis Method (CREAM)*. New York: Elsevier Science Inc.
		The method provides references that allow the information forming the technical basis and/or the data underpinning it to be obtained.	X	
		The method does not provide sufficient information to allow its technical basis and underpinning data to be accessed for review.		

Construct validity	Essential	**Attribute 2** **The Technical basis of the method (Theory)** The technical basis of the method is based upon, and does not contradict, a relevant body of scientific knowledge		CREAM
		The method operationalises a relevant model of human performance or system safety which has scientific acceptance.	X	Justification Model of human performance and human cognition is based on scientific approach, which is described in detail in the following chapters of the CREAM Handbook - Chapter 4: A Conceptual Framework. - Chapter 6: CREAM – Second Generation HRA Method. - Chapter 8: Qualitative Performance Prediction. This approach was considered appropriate at the time of the development. Since that time, developments (including those by the method's author) have advanced the understanding of cognition and error mechanisms.
		Elements of the method are inconsistent with an accepted scientific model of human performance or system safety.		

Construct validity	Essential	**Attribute 3** **The technical basis of the method (Data)** Where the technical basis of the method is based on a dataset, the source of the data/information and its relevance for application in the nuclear industry should be demonstrated.		CREAM
		The data underlying the method are largely based on observations of actual or simulated task performance in nuclear industry tasks.		Justification The data presented in Table 9 of the guidebook has made extensive use of the established data sources for proceduralised behaviours such as ***observation*** and ***execution***. While these CFPs are relatively well established, CFPs for ***interpretation*** and ***planning*** behaviour are mostly based on expert judgement. The values have been taken from a variety of sources…". These sources are largely nuclear related. After the paragraph follows a table with final values for the 13 different CFFs (note: the table contains 4 apparent mistakes in orders (decimal positions) without any other comments related to how these values were assigned to the CFFs. But there is no information or guidance on: how the data were combined; which specific data (from the list of sources) were used for which specific failure type or how the values were assigned to the failure types. The assignment of values is highly untraceable.
		The data underlying the method are based on expert judgement or observations of human performance for relevant tasks in a domain that is closely related to the nuclear industry e.g. other high hazard industries.	X	
		The data underlying the method are taken from tasks that are not related or relevant to nuclear industry tasks.		

Construct validity	Highly desirable	**Attribute 4** **Internal consistency of the method** The method demonstrates internal consistency between the technical basis, the error definition, the PSFs and the qualitative and quantitative method steps		CREAM
		The qualitative and quantitative component parts of the method are theoretically compatible and form a coherent consistent whole.	X	Justification There were observed no inconsistencies between the qualitative and/or quantitative parts of the method.
		There are theoretical inconsistencies between the qualitative and/or quantitative components of the method.		

Content Validity	Highly desirable	**Attribute 5** **Qualitative assessment** It is recognised good practice that HRA quantification is supported by qualitative analysis to develop an understanding of operator performance within the scenario that is being assessed. This attribute considers the extent to which the qualitative analysis stages of the HRA (e.g. task analysis and error identification) is directed or prescribed by the HRA method, beyond providing a set of performance shaping factors to be considered.		CREAM
		The method contains or prescribes a process for conducting qualitative assessment**.**	X	Justification CREAM requires a full decomposition of tasks corresponding to the CFFs, this process is described in the method documentation
		The method includes a general statement indicating that a qualitative assessment should be provided, e.g. by referring to the use of task analysis**.**		
		The method does not make any reference to qualitative analysis**.**		

Content validity	Essential	**Attribute 6** **Factors influencing human reliability considered by the method** The method should be quantitatively sensitive to a majority of accepted factors* (PSFs) that influence human reliability. *: There are pre-defined lists of PSFs/EPCs available throughout the literature and within HRA methods (typically 1st generation methods). This attribute does not seek to define such a list, so as to accommodate developments in human performance, system safety and accident analysis. The evaluation teams are asked to use professional judgment when considering this attribute. It is not expected that methods will be assigned a high rating (dark blue) if only a small number of factors are accommodated, but we do not prescribe an absolute number of factors that are required.		CREAM
		Sub-scale 1 **Adequacy of PSFs**.		CREAM
		The method requires qualitative assessment of a majority of accepted factors that affect human reliability (PSFs).	X	Justification (Reviewers should identify the PSFs that are included in the method and judge the adequacy of this set for the intended application) Instead of the term PSF, CREAM uses a term Common Performance Conditions (CPCs). The method contains 9 CPCs: 1. Adequacy of organisation. 2. Working conditions. 3. Adequacy of MMI and operational support. 4. Availability of procedures/plans. 5. Number of simultaneous goals. 6. Available time. 7. Time of day (circadian rhythm). 8. Adequacy of training and experience. 9. Crew collaboration quality This set of factors seems to be appropriate for the most of human failure events (HFEs) usually included in HRA/PSA as was practically verified in both HRA for NPP Dukovany and HRA for NPP Temelin. Factor "stress/stressors" is not considered explicitly, since it is presumed that most of the CPCs will act as stressors if they influence operators' performance in negative manner. It means that the factor stress/stressor is considered implicitly only, which seems to be sufficient.
		The method does not consider a majority set of factors that affect human reliability.		

Content validity	Essential	**Sub-scale 2** **Quantitative sensitivity**.		CREAM
		The method is quantitatively sensitive to the effect of each individual PSF considered qualitatively.	X	Justification Generally, the method is quantitatively sensitive to the effect of the most of considered CPCs. In some cases (e.g. crew collaboration quality), the method is sensitive only for the extreme states – very efficient (weighting factor 0,5) or deficient (weighting factor 2 or 5), but it doesn't distinguish between states efficient and inefficient (weighting factor 1 for both levels of the attribute). Situation is similar for CPC "Adequacy of MMI and operational support" and "Number of simultaneous goals".
		The method is not quantitatively sensitive to individual PSFs, but makes a single adjustment to the HEP based on the contribution of the overall combination of the PSFs considered.		
		The method is not quantitatively sensitive to PSFs.		
		Sub-scale 3 **Interaction between factors** Typically HRA methods adopt a linear multiplicative combination of PSFs. It is recognised that some PSFs may interact in other ways, e.g. a step change in the effect of one PSF once a threshold has been reached on a second PSF, or where the effect of the combination of two PSFs is far greater than multiplicative relationship would predict or where one PSF has a triggering effect on other PSFs in a causal chain.		CREAM
		Interactions between PSFs are accounted for on the basis of knowledge of the relationship between specific PSFs.	X	Justification Level of 4 CPCs (Working conditions, Number of goals, Available time, Crew collaboration quality) are adjusted after taking into account interactions between CPCs. This means that the derivation of the combined CPC score must take into account the way in which the CPCs are coupled or dependent. The rules for the adjustment are described in detail in the guidebook.
		Combinations of PSF effects are accounted for using a simple linear model.		
		Interactions between or combination of PSF effects are not considered by the method.		

Content validity	Essential	**Attribute 7** **Consideration of human error dependency** Modelling should include consideration of human error dependencies or common cause failures.		CREAM
		The method provides a procedure for identifying potential sources of dependence among Human Failure Events (HFEs) and/or sub-tasks of an HFE, and provides a method to derive conditional HEPs based on the systematic assessment of these sources of dependence.		Justification CREAM, does not consider dependency (neither negative nor positive) between two Human Failure Events and/or two steps of an HFE.
		The method identifies potential sources of dependence, but does not provide a process for linking these sources of dependence to a quantified model for deriving conditional HEPs.		
		The method does not address dependencies and common cause mechanisms among tasks and sub-tasks.	X	

Content validity	Essential	**Attribute 8** **Consideration of deviations and progressions in accident sequences** The method should provide a capability to accommodate: • Deviations from nominal accident scenarios due to: (A) Plant conditions: 1. Aleatory factors, such as sizes and locations of equipment failures and time sequences. 2. Complicating factors, such as coincident failures in control, instrumentation and support systems not normally modelled explicitly in PSA models. (B) Human failure scenarios due to crew organisational & operating practices that introduce opportunities to create new failure mechanisms. • Fault progressions including consequential faults and accident sequences encompassing Level 1 and Level 2 PSA which may involve extended time sequences and degraded operating environments should also be accommodated.	CREAM

Content validity	Essential	**Sub-scale 1** **Deviations**		CREAM
		The method provides for the qualitative and quantitative assessment of all the types of deviations in accident scenarios.		Justification The method does not deal with independent faults coincident in time explicitly, but the definitions of function failures and CPCs are so general, that it should not be a problem to take such cases into account. For example, CPC "Number of simultaneous goals" can be assessed as "more than capacity" in case of coincident faults, etc.
		The method provides for the qualitative assessment of human error during fault progressions, but does not provide for the derivation of HEPs in support of this assessment.	X	
		The method does not provide a means to deal with deviations in accident scenarios		
		Sub-scale 2 **Fault progression**		CREAM
		The method provides for the qualitative and quantitative assessment of human errors during fault progressions including level 1 to level 2 PSA fault progressions		Justification The method provides 13 failure modes called Cognitive Function Failures (CFFs) with their error probabilities. While these are appropriate to Level 1 actions, it is not clear that the CFFs or their values would apply (or be the only ones that apply) to fault progressions.
		The method provides for the qualitative assessment of human error during fault progressions, but does not provide for the derivation of HEPs in support of this assessment.		
		The method does not provide for the qualitative and quantitative assessment of human errors during fault progressions.	X	

Content validity	Highly desirable	**Attribute 9** **Consideration of cognitive error** The method should be sensitive to the factors that influence the diagnosis and decision making component of the response to an initiating event.		CREAM
		The method estimates the probability of cognitive error based on the assessment of a set of factors that are known to affect diagnosis and decision making performance	X	Justification The method takes into account 3 cognitive functions related to the cognitive errors: Observation, Interpretation and Planning. There are 8 failure types (with HEPs) connected with these Cognitive functions (see the first 8 items in the list). All of these failure types (and their HEPs) may be affected by any of the considered CPCs. This approach was considered largely state of the art at the time of the method's development.
		The method uses a simple model such as a time reliability curve as the primary factor for estimating the probability of cognitive error**.**		
		The method provides no way of estimating the likelihood of cognitive error**.**		

Content validity	Highly desirable	**Attribute 10** **Consideration of statistical uncertainty** The method should provide for statistical uncertainty analysis of derived human error probabilities.		CREAM
		The method derives uncertainty parameters from experience (either in-plant or from relevant simulator trials).		Justification The method provides nominal values and uncertainty bounds (lower and upper bound) for each of the cognitive failures.
		The method provides generic uncertainty parameters, e.g. standardised error factors.	X	
		The method provides no uncertainty parameters.		

Content validity	Desirable	**Attribute 11** **Consideration of organisational issues** The method should consider the impact of organisational issues including safety-culture factors (attitudes and behaviours), and organisational process factors (e.g. command and control structures, conflicts of interest, communication and decision making protocols on human reliability).		CREAM
		Sub-scale 1 **Safety-culture factors** (attitudes and behaviours).		CREAM
		The method provides an adequate quantitative method to adjust HEPs based on an assessment of safety culture/safety climate**.**		Justification The method does not provide any assessment of the safety culture.
		The method provides a qualitative means to assess safety culture/safety climate, but does not include a process to modify HEPs based on the assessment**.**		
		The method does not take into account safety culture factors**.**	X	
		Sub-scale 2 **Process factors** (e.g. command and control structures, communication and decision making protocols on human reliability).		CREAM
		The method provides a quantitative method to assess process factors	X	Justification CREAM includes some of the "process factors" mentioned above, and it provides a quantitative method to assess them. One of the CPCs included in the method is "Adequacy of organisation" which is described as follows: "The quality of the support and resources provided by the organisation for the task or work being performed. This includes communication system, Safety Management System, support for external activities, etc." Another CPC included in the method and related to this attribute is "Crew collaboration quality".
		The method provides a qualitative means to assess process factors, but does not include a process to modify HEPs based on the assessment**.**		
		The method does not take into account process factors**.**		

Empirical validity	Essential/Desirable	**Attribute 12** **Empirical validity** The method should demonstrate evidence of empirical validation exercises, peer review processes or community acceptance based on application and maturity.		CREAM
		Sub-scale1 **Statistical evidence**		CREAM
		The HEP estimates derived by the method have been shown to demonstrate good agreement with plant and /or simulator data for comparable tasks**.**		Justification The method has been included in the International HRA Empirical Study (Halden, 2007-2009) and US HRA Empirical Study (2009-2011). The method was assessed quite positively; the outcomes were in accordance with average outcomes of other methods. However due to the non-statistical treatment of the data generated by the international empirical study, it is not considered to provide evidence in relation to this attribute in this study.
		The HEP estimates derived by the method have been shown to demonstrate good agreement with HEP estimates produced by other HRA methods for the same or comparable tasks**.**		
		The method has failed to derive comparable HEP estimates in tests of empirical validity or has not been subject to such assessments**.**	X	
		Sub-scale 2 **Verification/Peer review**		CREAM
		The method has been subject to peer review by a team of recognised HRA experts, and the peer review comments have been incorporated to the development of the method**.**		Justification The method has been subject to peer review by a team of HRA experts during International HRA Empirical Study, however, the comments and recommendations have not been incorporated in the method.
		The method has been subject to peer review by a single, recognised HRA expert, and the comments have been incorporated to the development of the method**.**		
		The method has not been subject to independent peer review or the method has not been updated in response to peer review comments.	X	

Empirical validity	Essential/Desirable	**Sub-scale 3** **Application/Maturity**		CREAM
		The method has been extensively applied, internationally, for five or more years**.**		Justification In 2007, the method was used for NPP Dukovany HRA analysis and Temelin HRA analysis.
		The method has been applied to a limited number of HRAs**.**	X	
		The method has not yet been applied to a HRA**.**		

Reliability	Essential	**Attribute 13** **Computer models and software tools** If a method incorporates the use of a computer model or software tool to analyse a human action, A QA programme should be applied to ensure quality of the design and validity of the output.		CREAM
		A relevant, recognised/accepted international standard has been applied to the software design and verification of the computer based HRA method/tool**.**		Justification Not applicable. The method does not incorporate the use of any computer model or software.
		The design of the computer based HRA method/tool is based upon a documented QA process, which includes software verification**.**		
		There is no evidence that the design of the computer based HRA method/tool is based on a structured and validated software development or QA method that includes software verification**.**		

Reliability	Highly desirable	**Attribute 14** **Reliability and traceability** The method should provide consistent qualitative and quantitative information for comparable scenarios within analysts and between analysts for similar scenarios. The method should also provide sufficient information to facilitate tracing estimates back to input assumptions.		CREAM
		Sub-scale 1 **Within analyst consistency/reliability**		CREAM
		A formal comparison, amenable to statistical analysis, has been undertaken to demonstrate that the same HRA analyst provides consistent answers for analyses made at different times for the same scenario**.**		Justification We have no information about any formal or informal comparison focused on demonstration of consistency/reliability of the method.
		An informal comparison has been undertaken, which suggests good within analyst agreement for analyses made at different times**.**		
		There is no information available to suggest good within analyst agreement for analyses made at different times**.**	X	
		Sub-scale 2 **Between analyst consistency/reliability**		CREAM
		A formal comparison, amenable to statistical analysis, has been undertaken to demonstrate that different HRA analysts provide consistent answers for the same scenario**.**		Justification We have no information about any formal or informal comparison focused on demonstration of consistency/reliability of the method.
		An informal comparison has been undertaken, which suggests good between analyst agreement**.**		
		There is no information available to suggest good between analyst agreement**.**	X	

Reliability	Higly desirable	**Sub-scale 3** **Traceability**		CREAM
		The method provides a procedure to ensure easy, complete traceability of the estimates of human performance in the HRA, such that an independent reviewer could trace back HEPs to relevant assumptions, models and data cited in the method**.**	X	Justification Overall traceability of the method application is very good, as was stated in the report on the above mentioned International HRA Empirical Study.
		The HRA method itself does not provide a procedure for traceability, but there is sufficient information available about the method to facilitate traceability, and enable an independent reviewer to understand what was done**.**		
		There is insufficient information available to facilitate traceability**.**		

Usability	Highly desirable	**Attribute 15** **Definition of method scope** The scope of the method should be clearly defined.		CREAM
		The scope of the method is clearly defined in a user manual and/or technical basis document**.**	X	Justification The scope of the method is clearly defined in the guidebook.
		The scope of the method is described vaguely and some analyst judgement is required to determine its applicability to a particular human action/error**.**		Beside qualitative and quantitative performance predictions used for HRA purposes, CREAM also describes so called retrospective analyses (Chapter 7), which may be used for searching for causes of the accidents.
		The scope of the method is not defined**.**		

Usability	Highly desirable	**Attribute 16** **Qualitative outputs** The method should produce qualitative outputs that are useful to inform human factors and safety management improvements at the plant.		CREAM
		The method generates qualitative information to inform improvements to reduce the potential for human error that is explicitly related to each of the factors that are used in the method to derive an HEP.	X	Justification Based on the qualitative analysis, definitions of the identified Failure functions and Common performance conditions, it is very natural and easy to formulate recommendations leading to improvements which can reduce potential human errors. The reason is, as mentioned above, that CREAM uses similar concept not only for predictive analysis (HRA purposes), but also for retrospective analysis (described in Chapter 7), which is focused on searching for causes of the problems and accidents. After finding the causes, HRA expert can subsequently formulate recommendations, how the problems should be solved.
		The method generates qualitative information to inform improvements to reduce the potential for human error, but this is not explicitly linked to each of the factors used in the derivation of HEPs.		
		The method does not generate qualitative information to inform improvements to reduce the potential for human error.		

Usability	Highly desirable	**Attribute 17** **Qualitative uncertainty and quantitative conservatism** Methods should be able to reflect uncertainties related to qualitative information via conservatisms in the quantification process.		CREAM
		The method provides a mathematical procedure for adjusting the conservatism of the HEPs derived as a function of the level of certainty in the qualitative information collected during the assessment.		Justification The method does not provide any guidelines addressing this issue.
		The method provides a general caution on the need to adjust the conservatism of HEPs as a function of the level of certainty in the qualitative information collected, but does not provide a mathematical procedure for doing so.		
		The method does not address the issue of uncertainties in qualitative information and the impact of this on derived HEPs.	X	

Usability	Desirable	**Attribute 18** **Availability of user documentation** The method should be supported by a detailed user documentation e.g., manual or instructions, which describes how the method should be applied.		CREAM
		The method contains user documentation that provides a detailed step-by-step procedure for all steps in the derivation of an HEP.	X	Justification The method is described in the following publication: Erik Hollnagel: Cognitive Reliability and Error Analysis Method – CREAM (Elsevier, 1998), which describes step-by-step procedure for all steps in derivation of an HEP. More guidance should be provided for process of selection of the Cognitive Failure Functions, since the results are highly dependent on this step.
		The method contains user documentation that provides a high level description of how it is applied to derive HEPs, but not all elements of the method are detailed as step-by-step procedures.		
		The method provides only a high level description of its method of application and or data tables for the derivation of HEPs.		

Usability	Desirable	**Attribute 19** **Use of limiting values** The method should provide limiting values. (Relevant Good Practice documents discuss limiting values that are used in member countries).		CREAM
		The method provides limiting values and advice on their application.	X	Justification The lowest theoretically possible value which is allowed by CREAM is 2,56E-05 (=“action on wrong object” positively influenced (multiplied) by all applicable CPCs), which means that there is no problem with unrealistically low HEPs in CREAM application.
		The method provides advice on the need to limit claims on human performance but does not provide specific limiting values.		
		The method does not consider the use of limiting values.		

Usability	Indifferent/Essential	**Attribute 20** **Resources** A comparative estimate of the resources (time, cost, utility demands, level of specialist training required, level and type of knowledge required) needed to apply the method in comparison with other HRA methods.		CREAM
		The estimated cost of and time required for applying the HRA method is less than or comparable with that of other HRA methods**.**	X	Justification The resources required for applying the CREAM method are estimated to be comparable with other methods.
		The estimated cost of and time required for applying the HRA method is in excess of that required for application of other HRA methods**.**		

A2.9 Attribute Evaluations – FLIM

Desirable Attributes of HRA – Methods Evaluation Scale – FLIM

Instructions to assessors

Indicate your evaluation of how well a method meets the requirements of an attribute by placing a tick in the relevant coloured box.

Provide a succinct justification for the rating you have allocated in the text box labelled justification.

You should complete an evaluation for each attribute and each sub-scale of an attribute.

If an attribute is not relevant to the method you are evaluating then record this in the justification box with an explanation as to why the attribute is not relevant.

Your evaluation should be based on the application of the method exactly as it is described in the method's user documentation. If there are modifications to the process which improve its application then these can be noted in the justification box as potential improvements, but they should not be considered as part of the evaluation.

Within the method evaluation scale, a high rating (dark blue) indicates that the requirements of an attribute have been fully or largely met for the method's intended scope of application. An intermediate rating (medium blue) indicates that a method meets some, but not all of the requirements of the attribute. A low rating (light blue) indicates that the requirements of the attribute are not met or that no evidence is available in relation to the attribute for the method.

Construct validity	Essential	**Attribute 1** **Availability of information relating to the technical basis of the method** Information is provided on the technical basis of the method, in terms of its scientific underpinnings and data, in order to allow a judgement on the validity of the method to be made.		FLIM
		Comprehensive information on the technical basis and/or data underpinning the method is available and its application is discussed as part of the documentation of the method.		Justification There is little in the way of formal documentation of the method to allow judgement of the technical basis. The method is documented in NUREG 6144 but the description of the method is embedded in a large multi-volume report that details an evaluation of potential severe accidents during low power and shutdown operations at a US NPP. General descriptions are available in NUREG-1842 that includes an evaluation of the method. It is noted that the FLIM method is a development from the Success Likelihood Index Method (SLIM) documented in NUREG/CR-3518 (1988).
		The method provides references that allow the information forming the technical basis and/or the data underpinning it to be obtained.	X	
		The method does not provide sufficient information to allow its technical basis and underpinning data to be accessed for review.		

Construct validity	Essential	**Attribute 2** **The technical basis of the method (Theory)** The technical basis of the method is based upon, and does not contradict, a relevant body of scientific knowledge		FLIM
		The method operationalises a relevant model of human performance or system safety which has scientific acceptance.	X	Justification The underlying model assumes that relative importance weights and ratings of PSFs, obtained from expert judgement and related to a task, can be multiplied and then summed across PSFs to arrive at the Failure Likelihood Index (FLI). This approach is the same as that for SLIM, with the exception that it directly derives a "failure" likelihood index (FLI), rather than a success likelihood index (SLI) like SLIM/MAUD. The method broadly is consistent with the PSF type of HRA method.
		Elements of the method are inconsistent with an accepted scientific model of human performance or system safety.		

Construct validity	Essential	**Attribute 3** **The technical basis of the method (Data)** Where the technical basis of the method is based on a dataset, the source of the data/information and its relevance for application in the nuclear industry should be demonstrated.		FLIM
		The data underlying the method are largely based on observations of actual or simulated task performance in nuclear industry tasks.		Justification The basic data for this method are derived by using events with known HEPs as calibration events; the identification of appropriate calibration values for obtaining HEPs (data on similar events) is a critical aspect of the method. After the expert judges identify the PSFs relevant to the events they are quantifying, and weight and rate the PSFs in terms of their influence on an event, calibration values are identified and used in conjunction with the obtained FLI for the event, in order to derive the HEP. The PSF scales consist of expert judgement data on the qualitative rating of the elements contributing to a PSF.
		The data underlying the method are based on expert judgement or observations of human performance for relevant tasks in a domain that is closely related to the nuclear industry e.g. other high hazard industries.	X	
		The data underlying the method are taken from tasks that are not related or relevant to nuclear industry tasks.		

Construct validity	Highly desirable	**Attribute 4** **Internal consistency of the method** The method demonstrates internal consistency between the technical basis, the error definition, the PSFs and the qualitative and quantitative method steps.		FLIM
		The qualitative and quantitative component parts of the method are theoretically compatible and form a coherent consistent whole.	X	Justification Inclusion of PSF scaling guidance for the seven PSFs employed by the method supports the expert teams in considering each PSF comprehensively, including identification of particularly adverse or "error-forcing" performance conditions. In addition, since analysts could still use other PSFs as needed, FLIM's strengths are similar to those of SLIM/MAUD. Use of expert judges lends credence to the results, provided that the judges are qualified and familiar with the events being assessed.
		There are theoretical inconsistencies between the qualitative and/or quantitative components of the method.		

Content validity	Highly desirable	**Attribute 5** **Qualitative assessment** It is recognised good practice that HRA quantification is supported by qualitative analysis to develop an understanding of operator performance within the scenario that is being assessed. This attribute considers the extent to which the qualitative analysis stages of the HRA (e.g. task analysis and error identification) is directed or prescribed by the HRA method, beyond providing a set of performance shaping factors to be considered.		FLIM
		The method contains or prescribes a process for conducting qualitative assessment**.**	X	Justification The FLIM guidance in NUREG/CR-6144 for qualitative analysis consists of an overview of the steps in Section 8.3.3.1 and a table (Table 8-1, on page 8-22) providing a brief instruction as to what should be gathered with respect to each factor and the headings under which this information should be documented. Following these steps and instructions yields a qualitative analysis of the HFE (a description of the task and the aspects of the performance context that will support or detract from the reliability of the task).
		The method includes a general statement indicating that a qualitative assessment should be provided, e.g. by referring to the use of task analysis**.**		
		The method does not make any reference to qualitative analysis**.**		

Content validity	Essential	**Attribute 6** **Factors influencing human reliability considered by the method** The method should be quantitatively sensitive to a majority of accepted factors* (PSFs) that influence human reliability. *: There are pre-defined lists of PSFs/EPCs available throughout the literature and within HRA methods (typically 1st generation methods). This attribute does not seek to define such a list, so as to accommodate developments in human performance, system safety and accident analysis. The evaluation teams are asked to use professional judgment when considering this attribute. It is not expected that methods will be assigned a high rating (dark blue) if only a small number of factors are accommodated, but we do not prescribe an absolute number of factors that are required.	FLIM

Content validity	Essential	**Sub-scale 1** **Adequacy of PSFs**		FLIM
		The method requires qualitative assessment of a majority of accepted factors that affect human reliability (PSFs).	X	Justification (Reviewers should identify the PSFs that are included in the method and judge the adequacy of this set for the intended application) In principle, the method allows consideration of a wide range of PSFs and is flexible in terms of which PSFs are included. In practice, FLIM specifies a canonical set of 7 PSFs with rating scales that are representative of the main PSFs typically considered for Cat. C actions. FLIM can be extended for other actions involving additional PSFs. In this case, the rating scales would also need to be developed. The PSFs are: Adequacy of Time, Procedural Guidance, Training and Experience, Indications of Conditions (and HMI), Stress, Preceding and Concurrent Actions, and Complexity (requirements for coordination and communication, etc.)
		The method does not consider a majority set of factors that affect human reliability.		
		Sub-scale 2 **Quantitative sensitivity**		FLIM
		The method is quantitatively sensitive to the effect of each individual PSF considered qualitatively.	X	Justification The effects of each PSF are analysed individually.
		The method is not quantitatively sensitive to individual PSFs, but makes a single adjustment to the HEP based on the contribution of the overall combination of the PSFs considered.		
		The method is not quantitatively sensitive to PSFs.		

Content validity	Essential	**Sub-scale 3** **Interaction between factors** Typically HRA methods adopt a linear multiplicative combination of PSFs. It is recognised that some PSFs may interact in other ways, e.g. a step change in the effect of one PSF once a threshold has been reached on a second PSF, or where the effect of the combination of two PSFs is far greater than multiplicative relationship would predict or where one PSF has a triggering effect on other PSFs in a causal chain.		FLIM
		Interactions between PSFs are accounted for on the basis of knowledge of the relationship between specific PSFs.		Justification Use of a mathematical formula provides a traceable derivation of the obtained HEPs, as long as the basis for the weights and ratings of PSFs is thoroughly documented. The cumulative effect of multiple PSFs is the combination of their individual effects. These are additive in the FLI. Undesired effects from inappropriate mathematical combinations of PSFs may distort the results.
		Combinations of PSF effects are accounted for using a simple linear model.	X	
		Interactions between or combination of PSF effects are not considered by the method.		

Content validity	Essential	**Attribute 7** **Consideration of human error dependency** Modelling should include consideration of human error dependencies or common cause failures.		FLIM
		The method provides a procedure for identifying potential sources of dependence among Human Failure Events (HFEs) and/or sub-tasks of an HFE, and provides a method to derive conditional HEPs based on the systematic assessment of these sources of dependence.		Justification FLIM does not include a specific dependency model. A Cat. C HRA using FLIM expects that dependence analysis is performed externally and subsequent to the FLIM quantification.
		The method identifies potential sources of dependence, but does not provide a process for linking these sources of dependence to a quantified model for deriving conditional HEPs.	X	
		The method does not address dependencies and common cause mechanisms among tasks and sub-tasks.		

Content validity | Essential

Attribute 8
Consideration of deviations and progressions in accident sequences

FLIM

The method should provide a capability to accommodate:

- Deviations from nominal accident scenarios due to:
  - (A) Plant conditions:
    1. Aleatory factors, such as sizes and locations of equipment failures and time sequences.
    2. Complicating factors, such as coincident failures in control, instrumentation and support systems not normally modelled explicitly in PSA models.
  - (B) Human failure scenarios due to crew organisational & operating practices that introduce opportunities to create new failure mechanisms.
- Fault progressions including consequential faults and accident sequences encompassing Level 1 and Level 2 PSA which may involve extended time sequences and degraded operating environments should also be accommodated.

Sub-scale 1
Deviations

FLIM

The method provides for the qualitative and quantitative assessment of all the types of deviations in accident scenarios.		Justification The method allows adjustments of the PSF ratings for specific versions of sequences.
The method provides for the qualitative assessment of human error during fault progressions, but does not provide for the derivation of HEPs in support of this assessment.		With regard to the deviation types: 1A: Aleatory factors. FLIM can account for the effect of such factors but does not raise the issue of identifying such factors. (FLIM presumes the identification of HFEs (including the specification of the context of these HFEs) has been completed before its application.
The method does not provide a means to deal with deviations in accident scenarios	X	1B: Complicating factors. FLIM can account for such complications. Some of the PSF rating scales address some of these complicating factors. Crew organisation. FLIM does not allow such issues to be addressed.

Content validity	Essential	**Sub-scale 2** **Fault progression**		FLIM
		The method provides for the qualitative and quantitative assessment of human errors during fault progressions including level 1 to level 2 PSA fault progressions		Justification The applicability of FLIM for operator actions during (severe) accident conditions, as in Level 2 PSA, is contingent on two issues: • Whether the canonical 7 PSFs and their scales address the conditions associated with Level 2 PSA operator actions. The scales for the PSFs of the FLIM method do consider some of the main aspects of Level 2 actions particularly non-prescriptive procedural guidance e.g. SAMGs, extended time, and degraded operating environments. • Whether calibration HEPs could be obtained for Level 2 HFEs. This aspect is considered to be more problematic and prevents the assignment of a high rating. The overall judgement is that the method provides for the qualitative assessment of SOME of the factors that impact on human error during fault progressions.
		The method provides for the qualitative assessment of human error during fault progressions, but does not provide for the derivation of HEPs in support of this assessment.	X	
		The method does not provide for the qualitative and quantitative assessment of human errors during fault progressions.		

Content validity	Highly desirable	**Attribute 9** **Consideration of cognitive error** The method should be sensitive to the factors that influence the diagnosis and decision making component of the response to an initiating event.		FLIM
		The method estimates the probability of cognitive error based on the assessment of a set of factors that are known to affect diagnosis and decision making performance	X	Justification The FLIM method does not estimate a cognitive error as a distinct component from an implementation/execution error. The failure likelihood index accounts for both cognitive error as well as implementation/execution error; the resulting index and HEP is for the HFE as a whole. The PSFs and the scales used to rate the PSFs in FLIM explicitly address factors that are recognised as affecting diagnosis and decision-making performance. These include: – The availability of indications on which to decide that the task modelled by the HFE is required. – The presence of other indications that may mask this requirement. – The presence of other indications that may motivate the crew to select other tasks. – The familiarity (training) of the operators with respect to the cues associated with the task modelled by the HFE. – The adequacy of the time available for detection and decision (as well as implementation). – The context in terms of preceding and concurrent actions and issues that may distract the crew. – Procedural guidance for the decision, clarity of the guidance (and location of this guidance: step-by-step or in a continuously applicable instruction). – Training and experience of the crews with respect to the task modelled by the HFE.
		The method uses a simple model such as a time reliability curve as the primary factor for estimating the probability of cognitive error**.**		
		The method provides no way of estimating the likelihood of cognitive error**.**		

Content validity	Highly desirable	**Attribute 10** **Consideration of statistical uncertainty** The method should provide for statistical uncertainty analysis of derived human error probabilities.		FLIM
		The method derives uncertainty parameters from experience (either in-plant or from relevant simulator trials).		Justification The method explicitly allows for the assessment of uncertainties but these are based on judgement rather than actual data.
		The method provides generic uncertainty parameters, e.g. standardised error factors.	X	
		The method provides no uncertainty parameters.		

Content validity	Desirable	**Attribute 11** **Consideration of organisational issues** The method should consider the impact of organisational issues including safety-culture factors (attitudes and behaviours), and organisational process factors (e.g. command and control structures, conflicts of interest, communication and decision making protocols on human reliability).		FLIM
		Sub-scale 1 **Safety-culture factors** (attitudes and behaviours).		
		The method provides an adequate quantitative method to adjust HEPs based on an assessment of safety culture/safety climate**.**		Justification The canonical PSFs do not address safety culture.
		The method provides a qualitative means to assess safety culture/safety climate, but does not include a process to modify HEPs based on the assessment**.**		
		The method does not take into account safety culture factors**.**	X	
		Sub-scale 2 **Process factors** (e.g. command and control structures, communication and decision making protocols on human reliability).		FLIM
		The method provides a quantitative method to assess process factors	X	Justification Some aspects of process factors are addressed within the rating scales for the PSFs Training and Experience, and Complexity (requirements for coordination and communication, etc.). The limited consideration of process factors is the basis for the assignment of the intermediate rating. To the extent that they are considered, HEPs obtained with FLIM will reflect these factors
		The method provides a qualitative means to assess process factors, but does not include a process to modify HEPs based on the assessment**.**		
		The method does not take into account process factors**.**		

Empirical validity	Essential/Desirable	**Attribute 12** **Empirical validity** The method should demonstrate evidence of empirical validation exercises, peer review processes or community acceptance based on application and maturity.		FLIM
		Sub-scale1 **Statistical evidence**		
		The HEP estimates derived by the method have been shown to demonstrate good agreement with plant and /or simulator data for comparable tasks**.**		Justification The FLIM method has not been subject to an empirical validation exercise. The SLIM-MAUD method was subject to an empirical assessment (Zimolong, 1992) but the differences between FLIM and SLIM-MAUD do not allow the validation results for SLIM-MAUD to be readily transferred to the FLIM method. It is worth noting that for FLIM as well as SLIM-MAUD, the calibration step is problematic for an empirical validation of the probability values obtained with the method. Neither FLIM nor SLIM-MAUD provides any calibration values and there is a lack of calibration data (external to the methods) as an input to this step. The HEPs obtained with SLIM ultimately depend on the analysts' selection of calibration value; consequently, the empirical validity of the results depend on the analysts. This characteristic of the method furthermore would be expected to affect the repeatability of the method when used by different analysts. Zimolong B., Empirical evaluation of THERP, SLIM and ranking to estimate HEPs, Reliability Engineering & System Safety, 35(1), 1992, 1-11.
		The HEP estimates derived by the method have been shown to demonstrate good agreement with HEP estimates produced by other HRA methods for the same or comparable tasks**.**		
		The method has failed to derive comparable HEP estimates in tests of empirical validity or has not been subject to such assessments**.**	X	

Empirical validity	Essential/Desirable	**Sub-scale 2** **Verification/Peer review**		FLIM
		The method has been subject to peer review by a team of recognised HRA experts, and the peer review comments have been incorporated to the development of the method**.**		Justification The FLIM method has not been subject to an independent peer review. It was evaluated in the NRC's "Evaluation of Human Reliability Analysis Methods Against Good Practices" (NUREG-1842).
		The method has been subject to peer review by a single, recognised HRA expert, and the comments have been incorporated to the development of the method**.**		
		The method has not been subject to independent peer review or the method has not been updated in response to peer review comments.	X	
		Sub-scale 3 **Application/Maturity**		FLIM
		The method has been extensively applied, internationally, for five or more years**.**	X	Justification The method has been applied in several PRAs, both in Switzerland and the United States.
		The method has been applied to a limited number of HRAs**.**		
		The method has not yet been applied to a HRA**.**		

Reliability	Essential	**Attribute 13** **Computer models and software tools** If a method incorporates the use of a computer model or software tool to analyse a human action, A QA programme should be applied to ensure quality of the design and validity of the output.		FLIM
		A relevant, recognised/accepted international standard has been applied to the software design and verification of the computer based HRA method/tool**.**		Justification Not applicable.
		The design of the computer based HRA method/tool is based upon a documented QA process, which includes software verification**.**		
		There is no evidence that the design of the computer based HRA method/tool is based on a structured and validated software development or QA method that includes software verification**.**		

Reliability	Highly desirable	**Attribute 14** **Reliability and traceability** The method should provide consistent qualitative and quantitative information for comparable scenarios within analysts and between analysts for similar scenarios. The method should also provide sufficient information to facilitate tracing estimates back to input assumptions.		FLIM
		Sub-scale 1 **Within analyst consistency/reliability**		
		A formal comparison, amenable to statistical analysis, has been undertaken to demonstrate that the same HRA analyst provides consistent answers for analyses made at different times for the same scenario**.**		Justification No such evaluation has been mentioned in the available documentation. Note: FLIM has a number of features to support within-analyst and between-analyst consistency/reliability. Foremost among these are the PSF rating scales.
		An informal comparison has been undertaken, which suggests good within analyst agreement for analyses made at different times**.**		
		There is no information available to suggest good within analyst agreement for analyses made at different times**.**	X	
		Sub-scale 2 **Between analyst consistency/reliability**		FLIM
		A formal comparison, amenable to statistical analysis, has been undertaken to demonstrate that different HRA analysts provide consistent answers for the same scenario**.**		Justification No comparison has been performed. Note: FLIM has a number of features to support within-analyst and between-analyst consistency/reliability. Foremost among these are the PSF rating scales.
		An informal comparison has been undertaken, which suggests good between analyst agreement**.**		
		There is no information available to suggest good between analyst agreement**.**	X	

Reliability	Highly desirable	**Sub-scale 3** **Traceability**		FLIM
		The method provides a procedure to ensure easy, complete traceability of the estimates of human performance in the HRA, such that an independent reviewer could trace back HEPs to relevant assumptions, models and data cited in the method**.**		Justification The method makes clear each step in the analysis through the use of identified scales for the PSF ratings. Several parts of the analysis are very traceable but some aspects can be difficult to document and justify in practice.
		The HRA method itself does not provide a procedure for traceability, but there is sufficient information available about the method to facilitate traceability, and enable an independent reviewer to understand what was done**.**	X	
		There is insufficient information available to facilitate traceability**.**		

Usability	Highly desirable	**Attribute 15** **Definition of method scope** The scope of the method should be clearly defined.		FLIM
		The scope of the method is clearly defined in a user manual and/or technical basis document**.**	X	Justification The scope of the method is clearly defined in the available literature and is aimed principally toward post initiator human actions, although in principle it could be applied to other types of human action.
		The scope of the method is described vaguely and some analyst judgement is required to determine its applicability to a particular human action/error**.**		
		The scope of the method is not defined**.**		

Usability	Highly desirable	**Attribute 16** **Qualitative outputs** The method should produce qualitative outputs that are useful to inform human factors and safety management improvements at the plant		FLIM
		The method generates qualitative information to inform improvements to reduce the potential for human error that is explicitly related to each of the factors that are used in the method to derive an HEP.		Justification The ratings of each PSF identify effectively what areas of human performance (within the scope of the model) need to be improved, and the rating scale identifies what degree of change need to be made.
		The method generates qualitative information to inform improvements to reduce the potential for human error, but this is not explicitly linked to each of the factors used in the derivation of HEPs.	X	
		The method does not generate qualitative information to inform improvements to reduce the potential for human error.		

Usability	Highly desirable	**Attribute 17** **Qualitative uncertainty and quantitative conservatism** Methods should be able to reflect uncertainties related to qualitative information via conservatisms in the quantification process.		FLIM
		The method provides a mathematical procedure for adjusting the conservatism of the HEPs derived as a function of the level of certainty in the qualitative information collected during the assessment.		Justification FLIM does not address this issue. The method presumes that the analysis team has access to subject matter experts who know the plant conditions.
		The method provides a general caution on the need to adjust the conservatism of HEPs as a function of the level of certainty in the qualitative information collected, but does not provide a mathematical procedure for doing so.		
		The method does not address the issue of uncertainties in qualitative information and the impact of this on derived HEPs.	X	

Usability	Desirable	**Attribute 18** **Availability of user documentation** The method should be supported by a detailed user documentation e.g., manual or instructions, which describes how the method should be applied.		FLIM
		The method contains user documentation that provides a detailed step-by-step procedure for all steps in the derivation of an HEP.	X	Justification Full documentation of FLIM is publically available in NUREG-6144 which provides a step-by-step procedure for undertaking the qualitative and quantitative aspects of the analysis.
		The method contains user documentation that provides a high level description of how it is applied to derive HEPs, but not all elements of the method are detailed as step-by-step procedures.		
		The method provides only a high level description of its method of application and or data tables for the derivation of HEPs.		

Usability	Desirable	**Attribute 19** **Use of limiting values** The method should provide limiting values. (Relevant Good Practice documents discuss limiting values that are used in member countries).		FLIM
		The method provides limiting values and advice on their application.		Justification The use of the known HEP calibration values for similar events limits the values of HEPs that can be predicted. However, no limiting values are provided and the issue is not addressed by the method guidance.
		The method provides advice on the need to limit claims on human performance but does not provide specific limiting values.		
		The method does not consider the use of limiting values.	X	

Usability	Indifferent/Essential	**Attribute 20** **Resources** A comparative estimate of the resources (time, cost, utility demands, level of specialist training required, level and type of knowledge required) needed to apply the method in comparison with other HRA methods.		FLIM
		The estimated cost of and time required for applying the HRA method is less than or comparable with that of other HRA methods**.**		Justification This method is expected to require resources and effort typically the same as other PSF-based methods. FLIM presumes that the PSF ratings (and weights) are elicited from plant operators (and/or trainers?). This method is likely to require significant demands on utility resources, though it must be recognised that utility personnel (operators and trainers) should be part of any HRA study, to provide operating experience that is missed by analysts without such experience. The evaluation of the PSFs should be within the skill set of experienced HRA analysts, though training in the specific calibration points for the PSF ratings is suggested. The method requires some level of knowledge of the judgments to be made in assessing the PSFs. This can be accomplished by experience and knowledge or by training.
		The estimated cost of and time required for applying the HRA method is in excess of that required for application of other HRA methods**.**	X	

A2.10 Attribute Evaluations – HuRECA

Desirable Attributes of HRA – Methods Evaluation Scale – HuRECA

Instructions to assessors

Indicate your evaluation of how well a method meets the requirements of an attribute by placing a tick in the relevant coloured box.

Provide a succinct justification for the rating you have allocated in the text box labelled justification.

You should complete an evaluation for each attribute and each sub-scale of an attribute.

If an attribute is not relevant to the method you are evaluating then record this in the justification box with an explanation as to why the attribute is not relevant.

Your evaluation should be based on the application of the method exactly as it is described in the method's user documentation. If there are modifications to the process which improve its application, but which are not formally recorded in an update to the method's documentation, then these can be noted in the justification box as potential improvements, but they should not be considered as part of the evaluation.

Within the method evaluation scale, a high rating (dark blue) indicates that the requirements of an attribute have been fully or largely met for the method's intended scope of application. An intermediate rating (medium blue) indicates that a method meets some, but not all of the requirements of the attribute. A low rating (light blue) indicates that the requirements of the attribute are not met or that no evidence is available in relation to the attribute for the method.

Construct validity	Essential	**Attribute 1** **Availability of information relating to the technical basis of the method** Information is provided on the technical basis of the method, in terms of its scientific underpinnings and data, in order to allow a judgement on the validity of the method to be made.		HuRECA
		Comprehensive information on the technical basis and/or data underpinning the method is available and its application is discussed as part of the documentation of the method.	X	Justification The theoretical basis and underpinning data are available in technical report KAERI/TR-4385/2011 which is available currently only in Korean. There are no plans currently to produce a version of this report in English. It has not been possible to evaluate the quality of the information provided by the technical basis document, but the key contents of the technical report has been presented at the OECD HRA task meeting by the authors, and evaluation of the method against desirable attributes was done by the HRA task members.
		The method provides references that allow the information forming the technical basis and/or the data underpinning it to be obtained.		
		The method does not provide sufficient information to allow its technical basis and underpinning data to be accessed for review.		

Construct validity	Essential	**Attribute 2** **The technical basis of the method (Theory)** The technical basis of the method is based upon, and does not contradict, a relevant body of scientific knowledge		HuRECA
		The method operationalises a relevant model of human performance or system safety which has scientific acceptance.	X	Justification Based on the THERP/ASEP method, the method incorporates into the major constituents (i.e., PSFs) of the method the major design aspects of and human behaviour characteristics in a computer-based control room from the literature such as NUREG-0711, NUREG/CR-6634, NUREG/CR-6635, NUREG/CR-6690, and the simulator experiments under a computer-based mock-up environment.
		Elements of the method are inconsistent with an accepted scientific model of human performance or system safety.		

Construct validity	Essential	**Attribute 3** **The technical basis of the method (Data)** Where the technical basis of the method is based on a dataset, the source of the data/information and its relevance for application in the nuclear industry should be demonstrated.		HuRECA
		The data underlying the method are largely based on observations of actual or simulated task performance in nuclear industry tasks.		Justification The method uses expert judgement to derive weighting factors for reflecting the design level of computer-based design features such as computer-based procedures and soft controls.
		The data underlying the method are based on expert judgement or observations of human performance for relevant tasks in a domain that is closely related to the nuclear industry e.g. other high hazard industries.	X	
		The data underlying the method are taken from tasks that are not related or relevant to nuclear industry tasks.		

Construct validity	Highly desirable	**Attribute 4** **Internal consistency of the method** The method demonstrates internal consistency between the technical basis, the error definition, the PSFs and the qualitative and quantitative method steps		HuRECA
		The qualitative and quantitative component parts of the method are theoretically compatible and form a coherent consistent whole.	X	Justification There is internal consistency between the theoretical basis, data from simulator studies and task analysis, design-related PSFs, and quantitative calculation process.
		There are theoretical inconsistencies between the qualitative and/or quantitative components of the method.		

Content Validity	Highly desirable	**Attribute 5** **Qualitative assessment** It is recognised good practice that HRA quantification is supported by qualitative analysis to develop an understanding of operator performance within the scenario that is being assessed. This attribute considers the extent to which the qualitative analysis stages of the HRA (e.g. task analysis and error identification) is directed or prescribed by the HRA method, beyond providing a set of performance shaping factors to be considered.		HuRECA
		The method contains or prescribes a process for conducting qualitative assessment**.**		Justification Qualitative assessment including task analysis is conducted to determine the level of PSFs including the design-related PSFs of computer-based MCR, as well as to understand the scenario context.
		The method includes a general statement indicating that a qualitative assessment should be provided, e.g. by referring to the use of task analysis**.**	X	
		The method does not make any reference to qualitative analysis**.**		

Content validity	Essential	**Attribute 6** **Factors influencing human reliability considered by the method** The method should be quantitatively sensitive to a majority of accepted factors* (PSFs) that influence human reliability. *: There are pre-defined lists of PSFs/EPCs available throughout the literature and within HRA methods (typically 1st generation methods). This attribute does not seek to define such a list, so as to accommodate developments in human performance, system safety and accident analysis. The evaluation teams are asked to use professional judgment when considering this attribute. It is not expected that methods will be assigned a high rating (dark blue) if only a small number of factors are accommodated, but we do not prescribe an absolute number of factors that are required.		HuRECA
		Sub-scale 1 **Adequacy of PSFs**.		HuRECA
		The method requires qualitative assessment of a majority of accepted factors that affect human reliability (PSFs).	X	Justification (Reviewers should identify the PSFs that are included in the method and judge the adequacy of this set for the intended application) The method provides a majority of PSFs based on literature review of state of the art in HRA and ergonomics. In addition to that, the method provides more detailed attributes of computer-based design features such as computer-based procedures and soft controls to reflect the designed level of those features into estimating HEPs.
		The method does not consider a majority set of factors that affect human reliability.		

Content validity	Essential	**Sub-scale 2** **Quantitative sensitivity**		HuRECA
		The method is quantitatively sensitive to the effect of each individual PSF considered qualitatively.	X	Justification The method explicitly gives a different output for a different level of an individual PSF.
		The method is not quantitatively sensitive to individual PSFs, but makes a single adjustment to the HEP based on the contribution of the overall combination of the PSFs considered.		
		The method is not quantitatively sensitive to PSFs.		
		Sub-scale 3 **Interaction between factors** Typically HRA methods adopt a linear multiplicative combination of PSFs. It is recognised that some PSFs may interact in other ways, e.g. a step change in the effect of one PSF once a threshold has been reached on a second PSF, or where the effect of the combination of two PSFs is far greater than multiplicative relationship would predict or where one PSF has a triggering effect on other PSFs in a causal chain.		HuRECA
		Interactions between PSFs are accounted for on the basis of knowledge of the relationship between specific PSFs.		Justification Interactions between PSFs are not accounted for. Combinations of PSFs are modelled in a linear way.
		Combinations of PSF effects are accounted for using a simple linear model.	X	
		Interactions between or combination of PSF effects are not considered by the method.		

Content validity	Essential	**Attribute 7** **Consideration of human error dependency** Modelling should include consideration of human error dependencies or common cause failures.		HuRECA
		The method provides a procedure for identifying potential sources of dependence among Human Failure Events (HFEs) and/or sub-tasks of an HFE, and provides a method to derive conditional HEPs based on the systematic assessment of these sources of dependence.	X	Justification The method provides a method for considering dependency between HFEs by identifying potential sources of dependence. The THERP dependence model is used for determining level of dependency and conditional HEPs.
		The method identifies potential sources of dependence, but does not provide a process for linking these sources of dependence to a quantified model for deriving conditional HEPs.		
		The method does not address dependencies and common cause mechanisms among tasks and sub-tasks.		

Content validity	Essential	**Attribute 8** **Consideration of deviations and progressions in accident sequences** The method should provide a capability to accommodate: • Deviations from nominal accident scenarios due to: (A) Plant conditions: 1. Aleatory factors, such as sizes and locations of equipment failures and time sequences. 2. Complicating factors, such as coincident failures in control, instrumentation and support systems not normally modelled explicitly in PSA models. (B) Human failure scenarios due to crew organisational & operating practices that introduce opportunities to create new failure mechanisms. • Fault progressions including consequential faults and accident sequences encompassing Level 1 and Level 2 PSA which may involve extended time sequences and degraded operating environments should also be accommodated.	HuRECA

Content validity	Essential	**Sub-scale 1** **Deviations**		HuRECA
		The method provides for the qualitative and quantitative assessment of all the types of deviations in accident scenarios.		Justification The method does not support the qualitative assessment of deviations in accident sequences.
		The method provides for the qualitative assessment of human error during fault progressions, but does not provide for the derivation of HEPs in support of this assessment.		
		The method does not provide a means to deal with deviations in accident scenarios.	X	
		Sub-scale 2 **Fault progression**		HuRECA
		The method provides for the qualitative and quantitative assessment of human errors during fault progressions including level 1 to level 2 PSA fault progressions		Justification The method is not aimed for Level 2 PSA explicitly.
		The method provides for the qualitative assessment of human error during fault progressions, but does not provide for the derivation of HEPs in support of this assessment.		
		The method does not provide for the qualitative and quantitative assessment of human errors during fault progressions.	X	

Content validity	Highly desirable	**Attribute 9** **Consideration of cognitive error** The method should be sensitive to the factors that influence the diagnosis and decision making component of the response to an initiating event.		HuRECA
		The method estimates the probability of cognitive error based on the assessment of a set of factors that are known to affect diagnosis and decision making performance	X	Justification The method considers major factors affecting human decision reliability such as computerised procedures, information systems, training, decision-making burden. Especially the method considers specific design elements of computerised procedures in adjusting nominal diagnosis probability.
		The method uses a simple model such as a time reliability curve as the primary factor for estimating the probability of cognitive error**.**		
		The method provides no way of estimating the likelihood of cognitive error**.**		

Content validity	Highly desirable	**Attribute 10** **Consideration of statistical uncertainty** The method should provide for statistical uncertainty analysis of derived human error probabilities.		HuRECA
		The method derives uncertainty parameters from experience (either in-plant or from relevant simulator trials).		Justification The method provides generic uncertainty parameters, e.g. standardized error factors based on THERP/ASEP.
		The method provides generic uncertainty parameters, e.g. standardised error factors.	X	
		The method provides no uncertainty parameters.		

Content validity	Desirable	**Attribute 11** **Consideration of organisational issues** The method should consider the impact of organisational issues including safety-culture factors (attitudes and behaviours), and organisational process factors (e.g. command and control structures, conflicts of interest, communication and decision making protocols on human reliability).		HuRECA
		Sub-scale 1 **Safety-culture factors** (attitudes and behaviours).		HuRECA
		The method provides an adequate quantitative method to adjust HEPs based on an assessment of safety culture/safety climate**.**		Justification The method does not take into account safety culture factors.
		The method provides a qualitative means to assess safety culture/safety climate, but does not include a process to modify HEPs based on the assessment**.**		
		The method does not take into account safety culture factors**.**	X	
		Sub-scale 2 **Process factors** (e.g. command and control structures, communication and decision making protocols on human reliability).		HuRECA
		The method provides a quantitative method to assess process factors		Justification The method does not take into account process factors.
		The method provides a qualitative means to assess process factors, but does not include a process to modify HEPs based on the assessment**.**		
		The method does not take into account process factors**.**	X	

Empirical validity	Essential/Desirable	**Attribute 12** **Empirical validity** The method should demonstrate evidence of empirical validation exercises, peer review processes or community acceptance based on application and maturity.		HuRECA
		Sub-scale1 **Statistical evidence**		HuRECA
		The HEP estimates derived by the method have been shown to demonstrate good agreement with plant and /or simulator data for comparable tasks**.**		Justification The method has not been subject to such assessments yet.
		The HEP estimates derived by the method have been shown to demonstrate good agreement with HEP estimates produced by other HRA methods for the same or comparable tasks**.**		
		The method has failed to derive comparable HEP estimates in tests of empirical validity or has not been subject to such assessments**.**	X	
		Sub-scale 2 **Verification/Peer review**		HuRECA
		The method has been subject to peer review by a team of recognised HRA experts, and the peer review comments have been incorporated to the development of the method**.**		Justification The method has not been subject to independent peer review yet.
		The method has been subject to peer review by a single, recognised HRA expert, and the comments have been incorporated to the development of the method**.**		
		The method has not been subject to independent peer review or the method has not been updated in response to peer review comments.	X	

Empirical validiy	Essential/Desirable	**Sub-scale 3** **Application/Maturity**		HuRECA
		The method has been extensively applied, internationally, for five or more years**.**		Justification K-HRA (HuRECA's former method) has been used in several domestic HRAs, but HuRECA has not been applied to a HRA yet.
		The method has been applied to a limited number of HRAs**.**		
		The method has not yet been applied to a HRA**.**	X	

Reliability	Essential	**Attribute 13** **Computer models and software tools** If a method incorporates the use of a computer model or software tool to analyse a human action, A QA programme should be applied to ensure quality of the design and validity of the output.		HuRECA
		A relevant, recognised/accepted international standard has been applied to the software design and verification of the computer based HRA method/tool**.**		Justification The HuRECA tool was developed based on a documented QA process. The tool runs on Windows and iOS, but only Korean version is available.
		The design of the computer based HRA method/tool is based upon a documented QA process, which includes software verification**.**	X	
		There is no evidence that the design of the computer based HRA method/tool is based on a structured and validated software development or QA method that includes software verification**.**		

Reliability	Highly desirable	**Attribute 14** **Reliability and traceability** The method should provide consistent qualitative and quantitative information for comparable scenarios within analysts and between analysts for similar scenarios. The method should also provide sufficient information to facilitate tracing estimates back to input assumptions.		HuRECA
		Sub-scale 1 **Within analyst consistency/reliability**		HuRECA
		A formal comparison, amenable to statistical analysis, has been undertaken to demonstrate that the same HRA analyst provides consistent answers for analyses made at different times for the same scenario**.**		Justification This demonstration has not been conducted yet.
		An informal comparison has been undertaken, which suggests good within analyst agreement for analyses made at different times**.**		
		There is no information available to suggest good within analyst agreement for analyses made at different times**.**	X	
		Sub-scale 2 **Between analyst consistency/reliability**		HuRECA
		A formal comparison, amenable to statistical analysis, has been undertaken to demonstrate that different HRA analysts provide consistent answers for the same scenario**.**		Justification This demonstration has not been conducted yet.
		An informal comparison has been undertaken, which suggests good between analyst agreement**.**		
		There is no information available to suggest good between analyst agreement**.**	X	

Reliability	Highly desirable	**Sub-scale 3** **Traceability**		HuRECA
		The method provides a procedure to ensure easy, complete traceability of the estimates of human performance in the HRA, such that an independent reviewer could trace back HEPs to relevant assumptions, models and data cited in the method**.**	X	Justification The method provides a procedure and formal worksheet, so that it is easy to trace back HEPs to relevant assumptions, models, and data.
		The HRA method itself does not provide a procedure for traceability, but there is sufficient information available about the method to facilitate traceability, and enable an independent reviewer to understand what was done**.**		
		There is insufficient information available to facilitate traceability**.**		

Usability	Highly desirable	**Attribute 15** **Definition of method scope** The scope of the method should be clearly defined.		HuRECA
		The scope of the method is clearly defined in a user manual and/or technical basis document**.**	X	Justification The method is dedicated for HRA of pre- and post-initiator human actions at NPPs. The scope is clearly defined in a user manual.
		The scope of the method is described vaguely and some analyst judgement is required to determine its applicability to a particular human action/error**.**		
		The scope of the method is not defined**.**		

Usability	Highly desirable	**Attribute 16** **Qualitative outputs** The method should produce qualitative outputs that are useful to inform human factors and safety management improvements at the plant.		HuRECA
		The method generates qualitative information to inform improvements to reduce the potential for human error that is explicitly related to each of the factors that are used in the method to derive an HEP.	X	Justification HuRECA uses design-specific PSFs, which are used for deriving an HEP as well as for improving a design level.
		The method generates qualitative information to inform improvements to reduce the potential for human error, but this is not explicitly linked to each of the factors used in the derivation of HEPs.		
		The method does not generate qualitative information to inform improvements to reduce the potential for human error.		

Usability	Highly desirable	**Attribute 17** **Qualitative uncertainty and quantitative conservatism** Methods should be able to reflect uncertainties related to qualitative information via conservatisms in the quantification process.		HuRECA
		The method provides a mathematical procedure for adjusting the conservatism of the HEPs derived as a function of the level of certainty in the qualitative information collected during the assessment.		Justification The method does not address the issue of uncertainties in qualitative information in the derivation of HEPs.
		The method provides a general caution on the need to adjust the conservatism of HEPs as a function of the level of certainty in the qualitative information collected, but does not provide a mathematical procedure for doing so.		
		The method does not address the issue of uncertainties in qualitative information and the impact of this on derived HEPs.	X	

Usability	Desirable	**Attribute 18** **Availability of user documentation** The method should be supported by a detailed user documentation e.g., manual or instructions, which describes how the method should be applied.		HuRECA
		The method contains user documentation that provides a detailed step-by-step procedure for all steps in the derivation of an HEP.	X	Justification The method provides a step-by-step procedure for all required steps to get an HEP.
		The method contains user documentation that provides a high level description of how it is applied to derive HEPs, but not all elements of the method are detailed as step-by-step procedures.		
		The method provides only a high level description of its method of application and or data tables for the derivation of HEPs.		

Usability	Desirable	**Attribute 19** **Use of limiting values** The method should provide limiting values. (Relevant Good Practice documents discuss limiting values that are used in member countries).		HuRECA
		The method provides limiting values and advice on their application.	X	Justification The current version of HuRECA suggests 1.0E-6 for a joint HEP as a limiting value. It is under consideration to modify the value to 1.0E-5.
		The method provides advice on the need to limit claims on human performance but does not provide specific limiting values.		
		The method does not consider the use of limiting values.		

Usability	Indifferent/Essential	**Attribute 20** **Resources** A comparative estimate of the resources (time, cost, utility demands, level of specialist training required, level and type of knowledge required) needed to apply the method in comparison with other HRA methods.		HuRECA
		The estimated cost of and time required for applying the HRA method is less than or comparable with that of other HRA methods**.**	X	Justification The resources required for applying the method is estimated to be comparable with other 1st generation HRA methods.
		The estimated cost of and time required for applying the HRA method is in excess of that required for application of other HRA methods**.**		

APPENDIX 3 METHOD DEVELOPERS

Method Developers who provided initial input information to support the method evaluations were invited to review and provide comment on the final evaluation of the method they were involved in the development of. Rather than modify the final evaluations we report verbatim the method developer's response to the evaluation. We encourage readers to review this material to obtain additional information when making decisions about the appropriateness of methods for any particular application.

A3.1 Developer's Comments on Method Evaluation – Enhanced Bayesian THERP

Developers Comments on the evaluation are shown in bold text for each attribute where a comment has been raised.

Construct validity	Essential	**Attribute 1** **Availability of information relating to the technical basis of the method** Information is provided on the technical basis of the method, in terms of its scientific underpinnings and data, in order to allow a judgement on the validity of the method to be made.		Enhanced Bayesian THERP
		Comprehensive information on the technical basis and/or data underpinning the method is available and its application is discussed as part of the documentation of the method.		Justification There is little in the way of formal documentation of the method to allow judgement of the technical basis, though confidential descriptions are available in PRAs that have used the method. Conference papers and research reports providing an overview of the method are publicly available.
		The method provides references that allow the information forming the technical basis and/or the data underpinning it to be obtained.	X	
		The method does not provide sufficient information to allow its technical basis and underpinning data to be accessed for review.		

Construct validity	Essential	**Attribute 2** **The technical basis of the method (Theory)** The technical basis of the method is based upon, and does not contradict, a relevant body of scientific knowledge		Enhanced Bayesian THERP
		The method operationalises a relevant model of human performance or system safety which has scientific acceptance.	X	Justification The method broadly is consistent with the PSF type of HRA method. This is inferred from the PSFs used and their relationship with the underlying THERP T/RC.
		Elements of the method are inconsistent with an accepted scientific model of human performance or system safety.		

Construct validity	Essential	**Attribute 3** **The technical basis of the method (Data)** Where the technical basis of the method is based on a dataset, the source of the data/information and its relevance for application in the nuclear industry should be demonstrated.		Enhanced Bayesian THERP
		The data underlying the method are largely based on observations of actual or simulated task performance in nuclear industry tasks.		Justification The basic data for this method are derived from the THERP T/RC; however, there are unexplained deviations from the basic THERP T/RC . The effectiveness of the PSFs is largely judgemental on the part of the analysts, though guidance is provided from the early applications as exemplars for future analyses.
		The data underlying the method are based on expert judgement or observations of human performance for relevant tasks in a domain that is closely related to the nuclear industry e.g. other high hazard industries.	X	
		The data underlying the method are taken from tasks that are not related or relevant to nuclear industry tasks.		

Construct validity	Highly desirable	**Attribute 4** **Internal consistency of the method** The method demonstrates internal consistency between the technical basis, the error definition, the PSFs and the qualitative and quantitative method steps		Enhanced Bayesian THERP
		The qualitative and quantitative component parts of the method are theoretically compatible and form a coherent consistent whole.	X	Justification The combined use of the THERP T/RC and the PSFs as adjustments to its point estimates is a coherent approach, both qualitatively and quantitatively.
		There are theoretical inconsistencies between the qualitative and/or quantitative components of the method.		

Content Validity	Highly desirable	**Attribute 5** **Qualitative assessment** It is recognised good practice that HRA quantification is supported by qualitative analysis to develop an understanding of operator performance within the scenario that is being assessed. This attribute considers the extent to which the qualitative analysis stages of the HRA (e.g. task analysis and error identification) is directed or prescribed by the HRA method, beyond providing a set of performance shaping factors to be considered.	Enhanced Bayesian THERP	
		The method contains or prescribes a process for conducting qualitative assessment**.**		Justification The documentation generally refers to the use of typical HRA modelling methods.
		The method includes a general statement indicating that a qualitative assessment should be provided, e.g. by referring to the use of task analysis**.**	X	
		The method does not make any reference to qualitative analysis**.**		

Content validity	Essential	**Attribute 6** **Factors influencing human reliability considered by the method** The method should be quantitatively sensitive to a majority of accepted factors* (PSFs) that influence human reliability. *: There are pre-defined lists of PSFs/EPCs available throughout the literature and within HRA methods (typically 1st generation methods). This attribute does not seek to define such a list, so as to accommodate developments in human performance, system safety and accident analysis. The evaluation teams are asked to use professional judgment when considering this attribute. It is not expected that methods will be assigned a high rating (dark blue) if only a small number of factors are accommodated, but we do not prescribe an absolute number of factors that are required.	Enhanced Bayesian THERP

Content validity	Essential	**Sub-scale 1** **Adequacy of PSFs**.		Enhanced Bayesian THERP
		The method requires qualitative assessment of a majority of accepted factors that affect human reliability (PSFs).	X	Justification (Reviewers should identify the PSFs that are included in the method and judge the adequacy of this set for the intended application) The combination of PSFs and the TRC seem to cover most post-initiator human factors concerns. The PSFs used are: K1: Quality and relevance of procedures. K2: Quality and relevance of training. K3: Quality and relevance of feedback from process (MMI). K4: Mental load (stress) in the situation. K5: Need for coordination and communication.
		The method does not consider a majority set of factors that affect human reliability.		
		Sub-scale 2 **Quantitative sensitivity**		Enhanced Bayesian THERP
		The method is quantitatively sensitive to the effect of each individual PSF considered qualitatively.	X	Justification The effects of each PSF are analysed individually for their effect on the T/RC.
		The method is not quantitatively sensitive to individual PSFs, but makes a single adjustment to the HEP based on the contribution of the overall combination of the PSFs considered.		
		The method is not quantitatively sensitive to PSFs.		

Content validity	Essential	**Sub-scale 3** **Interaction between factors** Typically HRA methods adopt a linear multiplicative combination of PSFs. It is recognised that some PSFs may interact in other ways, e.g. a step change in the effect of one PSF once a threshold has been reached on a second PSF, or where the effect of the combination of two PSFs is far greater than multiplicative relationship would predict or where one PSF has a triggering effect on other PSFs in a causal chain.		Enhanced Bayesian THERP
		Interactions between PSFs are accounted for on the basis of knowledge of the relationship between specific PSFs.		Justification Combinations of the effects of PSFs are calculated using Bayesian mathematical equations, but the effects of individual PSFs are considered separately.
		Combinations of PSF effects are accounted for using a simple linear model.	X	
		Interactions between or combination of PSF effects are not considered by the method.		

Content validity	Essential	**Attribute 7** **Consideration of human error dependency** Modelling should include consideration of human error dependencies or common cause failures.		Enhanced Bayesian THERP
		The method provides a procedure for identifying potential sources of dependence among Human Failure Events (HFEs) and/or sub-tasks of an HFE, and provides a method to derive conditional HEPs based on the systematic assessment of these sources of dependence.	X	Justification Dependencies are identified both in the qualitative analysis phase and in the cut-set investigation phase. Full dependence is suggested for multiple operator actions in the same minimal cut set.
		The method identifies potential sources of dependence, but does not provide a process for linking these sources of dependence to a quantified model for deriving conditional HEPs.		
		The method does not address dependencies and common cause mechanisms among tasks and sub-tasks.		

<table>
<tr><td rowspan="5">Content validity</td><td rowspan="5">Essential</td><td colspan="3">Attribute 8
Consideration of deviations and progressions in accident sequences
The method should provide a capability to accommodate:
• Deviations from nominal accident scenarios due to:
(A) Plant conditions:
1. Aleatory factors, such as sizes and locations of equipment failures and time sequences.
2. Complicating factors, such as coincident failures in control, instrumentation and support systems not normally modelled explicitly in PSA models.
(B) Human failure scenarios due to crew organisational & operating practices that introduce opportunities to create new failure mechanisms.
• Fault progressions including consequential faults and accident sequences encompassing Level 1 and Level 2 PSA which may involve extended time sequences and degraded operating environments should also be accommodated.
Enhanced Bayesian THERP</td></tr>
<tr><td colspan="3">Sub-scale 1
Deviations
Enhanced Bayesian THERP</td></tr>
<tr><td>The method provides for the qualitative and quantitative assessment of all the types of deviations in accident scenarios.</td><td></td><td rowspan="3">Justification
The method does not provide any explicit means to identify deviations in accident sequences.
Disagree. The whole point of assessing PSFs is to assess the deviation from nominal.</td></tr>
<tr><td>The method provides for the qualitative assessment of human error during fault progressions, but does not provide for the derivation of HEPs in support of this assessment.</td><td></td></tr>
<tr><td>The method does not provide a means to deal with deviations in accident scenarios.</td><td>X</td></tr>
</table>

Content validity	Essential	**Sub-scale 2** **Fault progression**		Enhanced Bayesian THERP
		The method provides for the qualitative and quantitative assessment of human errors during fault progressions including level 1 to level 2 PSA fault progressions		Justification The underlying T/RC is based on the time operators have to respond to prevent core damage from occurring. In principle the same kinds of PSFs could be used for level 2 analyses. Conceptually the method could be used into level 2 events but there is no support for it at present. **"Core damage" is not essential. It is the available time window to keep the plant within whatever safety limits.** **Not only in principle but also in practice.** **Conceptually, the method always asks to define the decision making context on which the PSFs are assessed. If "fault progression" is part of the context, then it shall be taken into account. Level 2 vs. 1 are labels used in PRA, but for operators it's a matter of any stage during the accident sequence.**
		The method provides for the qualitative assessment of human error during fault progressions, but does not provide for the derivation of HEPs in support of this assessment.		
		The method does not provide for the qualitative and quantitative assessment of human errors during fault progressions.	X	

Content validity	Highly desirable	**Attribute 9** **Consideration of cognitive error** The method should be sensitive to the factors that influence the diagnosis and decision making component of the response to an initiating event.		Enhanced Bayesian THERP
		The method estimates the probability of cognitive error based on the assessment of a set of factors that are known to affect diagnosis and decision making performance	X	Justification The PSFs used are considered appropriate for the estimation of failures in cognition. The method is therefore more appropriate than just the use of the T/RC.
		The method uses a simple model such as a time reliability curve as the primary factor for estimating the probability of cognitive error**.**		
		The method provides no way of estimating the likelihood of cognitive error**.**		

Content validity	Highly desirable	**Attribute 10** **Consideration of statistical uncertainty** The method should provide for statistical uncertainty analysis of derived human error probabilities.		Enhanced Bayesian THERP
		The method derives uncertainty parameters from experience (either in-plant or from relevant simulator trials).		Justification The method explicitly allows for the assessment of uncertainties but these are based on judgement rather than actual data. **There are few examples in plant-specific PRAs, where the estimates have been updated with relevant simulator trials. Theoretically it would be easy incorporate experience, but due to practical constraints this is very seldom done.**
		The method provides generic uncertainty parameters, e.g. standardised error factors.	X	
		The method provides no uncertainty parameters.		

Content validity	Desirable	**Attribute 11** **Consideration of organisational issues** The method should consider the impact of organisational issues including safety-culture factors (attitudes and behaviours), and organisational process factors (e.g. command and control structures, conflicts of interest, communication and decision making protocols on human reliability).		Enhanced Bayesian THERP
		Sub-scale 1 **Safety-culture factors** (attitudes and behaviours).		Enhanced Bayesian THERP
		The method provides an adequate quantitative method to adjust HEPs based on an assessment of safety culture/safety climate**.**		Justification None of the PSFs used nor the T/RC **Explicitly** represent any safety culture factors.
		The method provides a qualitative means to assess safety culture/safety climate, but does not include a process to modify HEPs based on the assessment**.**		
		The method does not take into account safety culture factors**.**	X	

Content validity	Desirable	**Sub-scale 2** **Process factors** (e.g. command and control structures, communication and decision making protocols on human reliability).		Enhanced Bayesian THERP
		The method provides a quantitative method to assess process factors		Justification None of the PSFs used nor the T/RC represent any process factors, though one PSF requires consideration of the need for co-ordination and communication. There appears to be no assessment of their availability or quality. **Disagree. PSF for co-ordination and communication covers process factors.**
		The method provides a qualitative means to assess process factors, but does not include a process to modify HEPs based on the assessment**.**		
		The method does not take into account process factors**.**	X	

Empirical validity	Essential/Desirable	**Attribute 12** **Empirical validity** The method should demonstrate evidence of empirical validation exercises, peer review processes or community acceptance based on application and maturity.		Enhanced Bayesian THERP
		Sub-scale1 **Statistical evidence**		Enhanced Bayesian THERP
		The HEP estimates derived by the method have been shown to demonstrate good agreement with plant and /or simulator data for comparable tasks**.**		Justification It is understood that there are close agreements with data gathered in the International Benchmarking HRA study documented in NUREG/IA-0216 vol. 1,2,3, however due to the non-statistical treatment of the data generated by the international empirical study, it is not considered to provide evidence in relation to this attribute in this study. **Maybe that is the way things were reported in NUREG/IA-0216, but one can always look at the raw data from that study and see the rather good agreement.**
		The HEP estimates derived by the method have been shown to demonstrate good agreement with HEP estimates produced by other HRA methods for the same or comparable tasks**.**		
		The method has failed to derive comparable HEP estimates in tests of empirical validity or has not been subject to such assessments**.**	X	

Empirical validity	Essential/Desirable	**Sub-scale 2** **Verification/Peer review**		Enhanced Bayesian THERP
		The method has been subject to peer review by a team of recognised HRA experts, and the peer review comments have been incorporated to the development of the method**.**	X	Justification There have been several reviews of the method by regulatory bodies in Scandinavia which is the basis for the assignment of the high rating. The method is also part of the International Benchmarking Study and the Nordic/German HRA method comparison.
		The method has been subject to peer review by a single, recognised HRA expert, and the comments have been incorporated to the development of the method**.**		
		The method has not been subject to independent peer review or the method has not been updated in response to peer review comments.		
		Sub-scale 3 **Application/Maturity**		Enhanced Bayesian THERP
		The method has been extensively applied, internationally, for five or more years**.**		Justification The method has been applied in three PRAs and the International Benchmarking Study.
		The method has been applied to a limited number of HRAs**.**	X	
		The method has not yet been applied to a HRA**.**		

Reliability	Essential	**Attribute 13** **Computer models and software tools** If a method incorporates the use of a computer model or software tool to analyse a human action, A QA programme should be applied to ensure quality of the design and validity of the output.		Enhanced Bayesian THERP
		A relevant, recognised/accepted international standard has been applied to the software design and verification of the computer based HRA method/tool**.**		Justification N/A. The method uses off-the-shelf software (MS Excel).
		The design of the computer based HRA method/tool is based upon a documented QA process, which includes software verification**.**		
		There is no evidence that the design of the computer based HRA method/tool is based on a structured and validated software development or QA method that includes software verification**.**		
Reliability	Highly desirable	**Attribute 14** **Reliability and traceability** The method should provide consistent qualitative and quantitative information for comparable scenarios within analysts and between analysts for similar scenarios. The method should also provide sufficient information to facilitate tracing estimates back to input assumptions.		Enhanced Bayesian THERP
		Sub-scale 1 **Within analyst consistency/reliability**		Enhanced Bayesian THERP
		A formal comparison, amenable to statistical analysis, has been undertaken to demonstrate that the same HRA analyst provides consistent answers for analyses made at different times for the same scenario**.**		Justification No such evaluation has been mentioned in the available documentation.
		An informal comparison has been undertaken, which suggests good within analyst agreement for analyses made at different times**.**		
		There is no information available to suggest good within analyst agreement for analyses made at different times**.**	X	

Reliability	Highly desirable	**Sub-scale 2** **Between analyst consistency/reliability**		Enhanced Bayesian THERP
		A formal comparison, amenable to statistical analysis, has been undertaken to demonstrate that different HRA analysts provide consistent answers for the same scenario**.**		Justification It is noted that a team should undertake the assessment of PSFs and the Bayesian process combines their assessments. Hence the method can accommodate between-analyst differences. However there has been no formal test of between analyst reliability.
		An informal comparison has been undertaken, which suggests good between analyst agreement**.**		
		There is no information available to suggest good between analyst agreement**.**	X	
		Sub-scale 3 **Traceability**		Enhanced Bayesian THERP
		The method provides a procedure to ensure easy, complete traceability of the estimates of human performance in the HRA, such that an independent reviewer could trace back HEPs to relevant assumptions, models and data cited in the method**.**		Justification Whilst, the method makes clear each step in the analysis through the use of identified scales for the PSF ratings, the use of the T/RC and the results not all steps are provided in sufficient detail.
		The HRA method itself does not provide a procedure for traceability, but there is sufficient information available about the method to facilitate traceability, and enable an independent reviewer to understand what was done**.**	X	
		There is insufficient information available to facilitate traceability**.**		

Usability	Highly desirable	**Attribute 15** **Definition of method scope** The scope of the method should be clearly defined.		Enhanced Bayesian THERP
		The scope of the method is clearly defined in a user manual and/or technical basis document**.**	X	Justification The scope of the method is clearly defined in the available literature and is aimed at post-initiating event human actions.
		The scope of the method is described vaguely and some analyst judgement is required to determine its applicability to a particular human action/error**.**		
		The scope of the method is not defined**.**		

Usability	Highly desirable	**Attribute 16** **Qualitative outputs** The method should produce qualitative outputs that are useful to inform human factors and safety management improvements at the plant		Enhanced Bayesian THERP
		The method generates qualitative information to inform improvements to reduce the potential for human error that is explicitly related to each of the factors that are used in the method to derive an HEP.		Justification The ratings of each PSF identify effectively what areas of human performance (within the scope of the model) need to be improved, and the rating scale suggests what kinds of changes need to be made. However, no specific corrections are suggested.
		The method generates qualitative information to inform improvements to reduce the potential for human error, but this is not explicitly linked to each of the factors used in the derivation of HEPs.	X	
		The method does not generate qualitative information to inform improvements to reduce the potential for human error.		

Usability	Highly desirable	**Attribute 17** **Qualitative uncertainty and quantitative conservatism** Methods should be able to reflect uncertainties related to qualitative information via conservatisms in the quantification process.		Enhanced Bayesian THERP
		The method provides a mathematical procedure for adjusting the conservatism of the HEPs derived as a function of the level of certainty in the qualitative information collected during the assessment.		Justification The method provides limited guidance on how to accommodate uncertainties associated with input information. **The method allows different opinions about PSFs and treat these different views formally via the Bayesian approach.** **Regarding conservatism, there is e.g. a limit for low probabilities when the time window is long.**
		The method provides a general caution on the need to adjust the conservatism of HEPs as a function of the level of certainty in the qualitative information collected, but does not provide a mathematical procedure for doing so.	X	
		The method does not address the issue of uncertainties in qualitative information and the impact of this on derived HEPs.		

Usability	Highly desirable	**Attribute 18** **Availability of user documentation** The method should be supported by a detailed user documentation e.g., manual or instructions, which describes how the method should be applied.		Enhanced Bayesian THERP
		The method contains user documentation that provides a detailed step-by-step procedure for all steps in the derivation of an HEP.		Justification The method is described in a series of case studies in papers and reports. These are generally sufficient to understand the process of the method but are not explicitly a user manual.
		The method contains user documentation that provides a high level description of how it is applied to derive HEPs, but not all elements of the method are detailed as step-by-step procedures.	X	
		The method provides only a high level description of its method of application and or data tables for the derivation of HEPs.		

Usability	Desirable	**Attribute 19** **Use of limiting values** The method should provide limiting values. (Relevant Good Practice documents discuss limiting values that are used in member countries).		Enhanced Bayesian THERP
		The method provides limiting values and advice on their application.		Justification The use of the T/RC limits the values of HEPs that can be predicted. However, the use of multiple PSFs that are rated very good could lead to very low probabilities. There appears to be no prohibition or advice concerning this situation. **Maybe so, but in practice it has not been experienced that the overall multiplicative factor would have been "very low".**
		The method provides advice on the need to limit claims on human performance but does not provide specific limiting values.		
		The method does not consider the use of limiting values.	X	

Usability	Indifferent/Essential	**Attribute 20** **Resources** A comparative estimate of the resources (time, cost, utility demands, level of specialist training required, level and type of knowledge required) needed to apply the method in comparison with other HRA methods.		Enhanced Bayesian THERP
		The estimated cost of and time required for applying the HRA method is less than or comparable with that of other HRA methods**.**	X	Justification This method is expected to require resources and effort typically the same as other PSF-based methods. This method is not likely to require major demands on utility resources, though it must be recognized that utility personnel (operators and trainers) should be part of any HRA study, to provide operating experience that is missed by analysts without such experience. The evaluation of the PSFs should be within the skill set of experienced HRA analysts, though training in the specific anchor points for the PSF ratings is suggested. In most cases to date the method has been applied by its developers. Training to utility staff is normally provided during the application process. However, it is judged that any training to external users would not be onerous.
		The estimated cost of and time required for applying the HRA method is in excess of that required for application of other HRA methods**.**		

A3.2 Developer's comments on method evaluation – NARA

Developers Comments on the evaluation are shown in bold text for each attribute where a comment has been raised.

Construct validity	Essential	**Attribute 1** **Availability of information relating to the technical basis of the method** Information is provided on the technical basis of the method, in terms of its scientific underpinnings and data, in order to allow a judgement on the validity of the method to be made.		NARA
		Comprehensive information on the technical basis and/or data underpinning the method is available and its application is discussed as part of the documentation of the method.	X	Justification Comprehensive documentation (367 pages) on the NARA technique and its derivation is provided in the Technical Basis for NARA, a Method of Human Error Quantification, Issue 7, January 2012, Report CRA-BEGL-POW-J032. There is also a shorter User Manual. The Technical Basis contains all data used in derivation of the quantification aspects of the technique. The technical basis document also provides a discussion on the relationship between the NARA technique and human information processing models of human performance. The technical basis document details how the values (HEPs and EPC weights) are derived from data, the sources of all data being identified. The technical basis document is a proprietary document owned by EDF Nuclear Generation Limited and is not publically available.
		The method provides references that allow the information forming the technical basis and/or the data underpinning it to be obtained.		
		The method does not provide sufficient information to allow its technical basis and underpinning data to be accessed for review.		

Construct validity	Essential	**Attribute 2** **The technical basis of the method (Theory)** The technical basis of the method is based upon, and does not contradict, a relevant body of scientific knowledge		NARA
		The method operationalises a relevant model of human performance or system safety which has scientific acceptance.	X	Justification NARA is not a direct operationalisation of a single model of human performance or system safety. The NARA Technical Basis document provides a discussion of the technical basis of the method demonstrating how it relates to three error-related modelling traditions in Human Factors & Performance: Information Processing, the Skill, Rule and Knowledge-Based Behaviour model, and Reason's 'Slips, lapses and mistakes' model. The method therefore is not inconsistent with accepted scientific models, however, neither is it a direct operationalisation of relevant models.
		Elements of the method are inconsistent with an accepted scientific model of human performance or system safety.		

Construct validity	Essential	**Attribute 3** **The technical basis of the method (Data)** Where the technical basis of the method is based on a dataset, the source of the data/information and its relevance for application in the nuclear industry should be demonstrated.		NARA
		The data underlying the method are largely based on observations of actual or simulated task performance in nuclear industry tasks.	X	Justification The NARA Technical basis document identifies each data point used in the derivation of HEPs associated with each Generic Task Type (GTT) used in the method. Approximately 2/3 of these come from the nuclear industry with the remainder deriving from other industries. The technical basis document also identifies each of the data sources used in establishing the maximum affect associated with each Error Producing Condition (NARA term for PSF). The majority of data used to derive numeric values for this aspect of the method are from laboratory experiments often using simple tasks that form component parts of nuclear industry tasks.
		The data underlying the method are based on expert judgement or observations of human performance for relevant tasks in a domain that is closely related to the nuclear industry e.g. other high hazard industries.		
		The data underlying the method are taken from tasks that are not related or relevant to nuclear industry tasks.		

Construct validity	Highly desirable	**Attribute 4** **Internal consistency of the method** The method demonstrates internal consistency between the technical basis, the error definition, the PSFs and the qualitative and quantitative method steps		NARA
		The qualitative and quantitative component parts of the method are theoretically compatible and form a coherent consistent whole.	X	Justification NARA demonstrates internal consistency between the quantification procedures and the theoretical basis which is largely founded on an information processing model. The quantification processes themselves are internally consistent with HEPs being assigned on the basis of generic task characteristics and these being modified on the basis of performance shaping factors including extended time factors.
		There are theoretical inconsistencies between the qualitative and/or quantitative components of the method.		

Content Validity	Highly desirable	**Attribute 5** **Qualitative assessment** It is recognised good practice that HRA quantification is supported by qualitative analysis to develop an understanding of operator performance within the scenario that is being assessed. This attribute considers the extent to which the qualitative analysis stages of the HRA (e.g. task analysis and error identification) is directed or prescribed by the HRA method, beyond providing a set of performance shaping factors to be considered.		NARA
		The method contains or prescribes a process for conducting qualitative assessment**.**		Justification The NARA user manual identifies that a task and error analysis should be conducted wherever possible to underpin the quantitative analysis provided by NARA. The Manual also identifies that such qualitative analysis is outside of the scope of the manual. **A comment. As you know, NARA is proprietary, and takes place within the context of a mature PSA and Human Factors environment. NARA does not define the task analysis etc. because that is defined elsewhere in the PSA/HFA 'infrastructure'. NARA has recently been formalized by EdF NGL in its Human Factors Integration Manual (HFIM). The manual has recently been applied to a pilot study based on a selection of operator actions claimed following boiler tube leak faults.**
		The method includes a general statement indicating that a qualitative assessment should be provided, e.g. by referring to the use of task analysis**.**	X	
		The method does not make any reference to qualitative analysis**.**		

Content validity	Essential	**Attribute 6** **Factors influencing human reliability considered by the method** The method should be quantitatively sensitive to a majority of accepted factors* (PSFs) that influence human reliability. *: There are pre-defined lists of PSFs/EPCs available throughout the literature and within HRA methods (typically 1st generation methods). This attribute does not seek to define such a list, so as to accommodate developments in human performance, system safety and accident analysis. The evaluation teams are asked to use professional judgment when considering this attribute. It is not expected that methods will be assigned a high rating (dark blue) if only a small number of factors are accommodated, but we do not prescribe an absolute number of factors that are required.	NARA

Content validity	Essential	**Sub-scale 1** **Adequacy of PSFs**		NARA
		The method requires qualitative assessment of a majority of accepted factors that affect human reliability (PSFs).	X	Justification (Reviewers should identify the PSFs that are included in the method and judge the adequacy of this set for the intended application) NARA has 18 PSFs (called error Producing Conditions, EPCs) listed below that should be considered during application: 1. Poor, ambiguous or ill-matched system feedback. 2. Unfamiliarity. 3. A need to unlearn a technique and apply one which requires the application of an opposing philosophy. 4. Time pressure. 5. Low signal to noise ratio. 6. A conflict between immediate and long-term objectives. 7. No obvious means of reversing an unintended action, 8. A means of suppressing or over-riding information or features which is too easily accessible. 9. Operator inexperience. 10. Cognitive overload, particularly one caused by simultaneous presentation of non-redundant information. 11. No obvious way of keeping track of progress during an activity. 12. Shortfalls in the quality of information conveyed by procedures. 13. Difficulties caused by poor shift hand-over practices and/or team co-ordination problems or friction between team members. 14. An incentive to use other more dangerous procedures to achieve long-term objectives. 15. Poor environment. 16. High emotional stress and effects of ill health. 17. Low workforce morale or adverse organisational environment. 18. Operator under-load/boredom The set of PSFs overlaps with those used in other HRA methods of this type and is consistent with relevant good practice as outlined e.g. in the USNRC Good Practices for implementing HRA guidance.
		The method does not consider a majority set of factors that affect human reliability.		

Content validity	Essential	**Sub-scale 2** **Quantitative sensitivity**		NARA
		The method is quantitatively sensitive to the effect of each individual PSF considered qualitatively.	X	Justification Each PSF (EPC) has its own independent quantitative weighting (effect on performance reliability).
		The method is not quantitatively sensitive to individual PSFs, but makes a single adjustment to the HEP based on the contribution of the overall combination of the PSFs considered.		
		The method is not quantitatively sensitive to PSFs.		
		Sub-scale 3 **Interaction between factors** Typically HRA methods adopt a linear multiplicative combination of PSFs. It is recognised that some PSFs may interact in other ways, e.g. a step change in the effect of one PSF once a threshold has been reached on a second PSF, or where the effect of the combination of two PSFs is far greater than multiplicative relationship would predict or where one PSF has a triggering effect on other PSFs in a causal chain.		NARA
		Interactions between PSFs are accounted for on the basis of knowledge of the relationship between specific PSFs.		Justification No PSF (EPC) interaction effects are considered. NARA uses a simple linear model.
		Combinations of PSF effects are accounted for using a simple linear model.	X	
		Interactions between or combination of PSF effects are not considered by the method.		

Content validity	Essential	**Attribute 7** **Consideration of human error dependency** Modelling should include consideration of human error dependencies or common cause failures.		NARA
		The method provides a procedure for identifying potential sources of dependence among Human Failure Events (HFEs) and/or sub-tasks of an HFE, and provides a method to derive conditional HEPs based on the systematic assessment of these sources of dependence.		Justification The method does not provide a procedure for dealing with dependency. It does however provide guidance on methods that can be used to address the issue, but these are not an integral part of the method **The NARA technique refers to and explains the use of High and Low Tier Contingent Actions as a dependency approach, as well as use of the THERP dependency model, and HPLVs. Perhaps this isn't enough to make it green [high rating]?**
		The method identifies potential sources of dependence, but does not provide a process for linking these sources of dependence to a quantified model for deriving conditional HEPs.	X	
		The method does not address dependencies and common cause mechanisms among tasks and sub-tasks.		

Content validity	Essential	**Attribute 8** **Consideration of deviations and progressions in accident sequences** The method should provide a capability to accommodate: • Deviations from nominal accident scenarios due to: (A) Plant conditions: 1. Aleatory factors, such as sizes and locations of equipment failures and time sequences. 2. Complicating factors, such as coincident failures in control, instrumentation and support systems not normally modelled explicitly in PSA models. (B) Human failure scenarios due to crew organisational & operating practices that introduce opportunities to create new failure mechanisms. • Fault progressions including consequential faults and accident sequences encompassing Level 1 and Level 2 PSA which may involve extended time sequences and degraded operating environments should also be accommodated.	NARA

Content validity	Essential	**Sub-scale 1** **Deviations**		NARA
		The method provides for the qualitative and quantitative assessment of all the types of deviations in accident scenarios.		Justification NARA does not provide the qualitative assessment tools required to model deviations in accident sequences.
		The method provides for the qualitative assessment of human error during fault progressions, but does not provide for the derivation of HEPs in support of this assessment.		
		The method does not provide a means to deal with deviations in accident scenarios	X	
		Sub-scale 2 **Fault progression**		NARA
		The method provides for the qualitative and quantitative assessment of human errors during fault progressions including level 1 to level 2 PSA fault progressions		Justification NARA does not provide the qualitative assessment tools required to model progressions in accident sequences. NARA contains a method for dealing with extended time factors where these may have a positive impact on human performance, this aspect of the method may provide some benefit for considering fault progressions. **I have a general question about fault progression (i.e. not specific to NARA). What if another qualitative approach is used to identify the fault progression, and then NARA (or THERP or whatever) is used to quantify it?**
		The method provides for the qualitative assessment of human error during fault progressions, but does not provide for the derivation of HEPs in support of this assessment.		
		The method does not provide for the qualitative and quantitative assessment of human errors during fault progressions.	X	

Content validity	Highly desirable	**Attribute 9** **Consideration of cognitive error** The method should be sensitive to the factors that influence the diagnosis and decision making component of the response to an initiating event.		NARA
		The method estimates the probability of cognitive error based on the assessment of a set of factors that are known to affect diagnosis and decision making performance	X	Justification NARA contains three GTTs relevant to cognitive error which map onto Rasmussen's Skill, Rule, Knowledge framework. A number of EPCs that affect decision-making and diagnosis e.g. cognitive overload, low signal to noise ratio are considered by the method.
		The method uses a simple model such as a time reliability curve as the primary factor for estimating the probability of cognitive error**.**		
		The method provides no way of estimating the likelihood of cognitive error**.**		

Content validity	Highly desirable	**Attribute 10** **Consideration of statistical uncertainty** The method should provide for statistical uncertainty analysis of derived human error probabilities.		NARA
		The method derives uncertainty parameters from experience (either in-plant or from relevant simulator trials).		Justification The HEPs associated with GTTs have uncertainty bounds (5-95%) which are statistically-derived based on the number of data points (and their range) used to derive the GTT.
		The method provides generic uncertainty parameters, e.g. standardised error factors.	X	
		The method provides no uncertainty parameters.		

<table>
<tr><td rowspan="8">Content validity</td><td rowspan="8">Desirable</td><td colspan="3">Attribute 11
Consideration of organisational issues
The method should consider the impact of organisational issues including safety-culture factors (attitudes and behaviours), and organisational process factors (e.g. command and control structures, conflicts of interest, communication and decision making protocols on human reliability).</td><td>NARA</td></tr>
<tr><td colspan="3">Sub-scale 1
Safety-culture factors (attitudes and behaviours).</td><td>NARA</td></tr>
<tr><td>The method provides an adequate quantitative method to adjust HEPs based on an assessment of safety culture/safety climate.</td><td>X</td><td rowspan="3">Justification
NARA considers a number of EPCs that relate to some aspects of safety culture e.g. a conflict between immediate and long term objectives, an incentive to use other, more dangerous procedures to achieve long-term objectives, low work force morale or adverse organisational environment. Whilst the EPCs may not address all of the components of safety culture, they provide some basic relevant factors to be addressed quantitatively.</td></tr>
<tr><td>The method provides a qualitative means to assess safety culture/safety climate, but does not include a process to modify HEPs based on the assessment.</td><td></td></tr>
<tr><td>The method does not take into account safety culture factors.</td><td></td></tr>
<tr><td colspan="3">Sub-scale 2
Process factors
(e.g. command and control structures, communication and decision making protocols on human reliability).</td><td>NARA</td></tr>
<tr><td>The method provides a quantitative method to assess process factors</td><td>X</td><td rowspan="3">Justification
The method provides a number of EPCs that relate to organisational process factors these include:
• Difficulties caused by poor shift hand-over practices and/or team co-ordination problems or friction between team members.
• Incentives to use more dangerous procedures.
• Low workforce morale or adverse organisational environment.
However the whole set of organisational process factors are not considered to be addressed by the method.</td></tr>
<tr><td>The method provides a qualitative means to assess process factors, but does not include a process to modify HEPs based on the assessment.</td><td></td></tr>
<tr><td>The method does not take into account process factors.</td><td></td></tr>
</table>

Empirical validity	Essential/Desirable	**Attribute 12** **Empirical validity** The method should demonstrate evidence of empirical validation exercises, peer review processes or community acceptance based on application and maturity.		NARA
		Sub-scale1 **Statistical evidence**		NARA
		The HEP estimates derived by the method have been shown to demonstrate good agreement with plant and /or simulator data for comparable tasks**.**		Justification NARA has not been subjected to empirical validations. **No validations, but it is closely modelled on HEART, which has been validated several times (and HEART did quite well in the recent Benchmark if I recall). I do not think you can change the rating, but perhaps a comment could be made on the justification statement to this effect.**
		The HEP estimates derived by the method have been shown to demonstrate good agreement with HEP estimates produced by other HRA methods for the same or comparable tasks**.**		
		The method has failed to derive comparable HEP estimates in tests of empirical validity or has not been subject to such assessments**.**	X	
		Sub-scale 2 **Verification/Peer review**		NARA
		The method has been subject to peer review by a team of recognised HRA experts, and the peer review comments have been incorporated to the development of the method**.**	X	Justification NARA has been subject to two independent international Peer Reviews with six HRA experts who gave anonymous comments which the NARA development team had to respond to and resolve to the satisfaction of the experts and the regulator who commissioned the reviews. The reviews resulted in a number of changes to the method.
		The method has been subject to peer review by a single, recognised HRA expert, and the comments have been incorporated to the development of the method**.**		
		The method has not been subject to independent peer review or the method has not been updated in response to peer review comments.		

Empirical validity	Essential/Desirable	**Sub-scale 3** **Application/Maturity**		NARA
		The method has been extensively applied, internationally, for five or more years**.**		Justification NARA has been applied to only a limited number of HRAs in the UK having only recently replaced HEART as an identified method for conducting HRA within EDF NGL. NARA was used as a quantification tool in the US in the Yucca Mountain HRA. **NARA is being used for all newly identified operator interventions within EDF NGL PSAs, and as mentioned, the need to use NARA has been formalised in the EDF NGL HFIM.**
		The method has been applied to a limited number of HRAs**.**	X	
		The method has not yet been applied to a HRA**.**		

Reliability	Essential	**Attribute 13** **Computer models and software tools** If a method incorporates the use of a computer model or software tool to analyse a human action, A QA programme should be applied to ensure quality of the design and validity of the output.		NARA
		A relevant, recognised/accepted international standard has been applied to the software design and verification of the computer based HRA method/tool**.**		Justification Not Applicable. NARA has not been developed as a software tool.
		The design of the computer based HRA method/tool is based upon a documented QA process, which includes software verification**.**		
		There is no evidence that the design of the computer based HRA method/tool is based on a structured and validated software development or QA method that includes software verification**.**		

Reliability	Highly desirable	**Attribute 14** **Reliability and traceability** The method should provide consistent qualitative and quantitative information for comparable scenarios within analysts and between analysts for similar scenarios. The method should also provide sufficient information to facilitate tracing estimates back to input assumptions.		NARA
		Sub-scale 1 **Within analyst consistency/reliability**		NARA
		A formal comparison, amenable to statistical analysis, has been undertaken to demonstrate that the same HRA analyst provides consistent answers for analyses made at different times for the same scenario**.**		Justification NARA has not been subject to any tests of within user reliability.
		An informal comparison has been undertaken, which suggests good within analyst agreement for analyses made at different times**.**		
		There is no information available to suggest good within analyst agreement for analyses made at different times**.**	X	
		Sub-scale 2 **Between analyst consistency/reliability**		NARA
		A formal comparison, amenable to statistical analysis, has been undertaken to demonstrate that different HRA analysts provide consistent answers for the same scenario**.**		Justification NARA has not been subject to any tests of between user reliability.
		An informal comparison has been undertaken, which suggests good between analyst agreement**.**		
		There is no information available to suggest good between analyst agreement**.**	X	

Reliability	Highly desirable	**Sub-scale 3** **Traceability**		NARA
		The method provides a procedure to ensure easy, complete traceability of the estimates of human performance in the HRA, such that an independent reviewer could trace back HEPs to relevant assumptions, models and data cited in the method**.**	X	Justification There is a strong emphasis on documentation and proforma for recording all assumptions, calculations, etc. This is dealt with in the User Manual and examples are given of how to document NARA usage. This is also heavily emphasised in the NARA training.
		The HRA method itself does not provide a procedure for traceability, but there is sufficient information available about the method to facilitate traceability, and enable an independent reviewer to understand what was done**.**		
		There is insufficient information available to facilitate traceability**.**		
Usability	Highly desirable	**Attribute 15** **Definition of method scope** The scope of the method should be clearly defined.		NARA
		The scope of the method is clearly defined in a user manual and/or technical basis document**.**		Justification The scope of the method is identified in the User Manual as to address the HEP assessment needs for a typical PSA, although what is meant by a typical PSA is not defined. The user manual also identifies that NARA is developed to consider fault sequences covering extended timescales and the need to cover dependencies. It is also identified that a module to cover errors of commission was produced but that this was considered tentative and as a result was not incorporated into the final technique. The user manual provides worked examples illustrating how NARA may be applied to quantification of pre-fault (maintenance) errors and post fault errors covering both diagnosis and action components including control room and on plant actions. **I am a bit puzzled by this one, as I think the scope is pretty clear? A typical NPP PSA is clearly defined in the United Kingdom under licensing requirements. The EOC module remains dormant. NARA quantifies the rest including pre-trip, post-trip, and ETFs, and the manual gives examples. I couldn't see what wasn't clear?**
		The scope of the method is described vaguely and some analyst judgement is required to determine its applicability to a particular human action/error**.**	X	
		The scope of the method is not defined**.**		

Usability	Highly desirable	**Attribute 16** **Qualitative outputs** The method should produce qualitative outputs that are useful to inform human factors and safety management improvements at the plant		NARA
		The method generates qualitative information to inform improvements to reduce the potential for human error that is explicitly related to each of the factors that are used in the method to derive an HEP.	X	Justification The output from a NARA analysis identifies the EPCs that have been used in deriving the HEP value. Information is available in the discussion of EPCs and their anchor values which can be used to derive recommendations for plant and operational improvements.
		The method generates qualitative information to inform improvements to reduce the potential for human error, but this is not explicitly linked to each of the factors used in the derivation of HEPs.		
		The method does not generate qualitative information to inform improvements to reduce the potential for human error.		

Usability	Highly desirable	**Attribute 17** **Qualitative uncertainty and quantitative conservatism** Methods should be able to reflect uncertainties related to qualitative information via conservatisms in the quantification process.		NARA
		The method provides a mathematical procedure for adjusting the conservatism of the HEPs derived as a function of the level of certainty in the qualitative information collected during the assessment.		Justification The User Manual does not explicitly address the issue of uncertainty related to qualitative information, however, NARA does provide a mechanism, the assessed proportion of affect, by which such uncertainties could be taken into account when deriving HEPs.
		The method provides a general caution on the need to adjust the conservatism of HEPs as a function of the level of certainty in the qualitative information collected, but does not provide a mathematical procedure for doing so.	X	
		The method does not address the issue of uncertainties in qualitative information and the impact of this on derived HEPs.		

Usability	Desirable	**Attribute 18** **Availability of user documentation** The method should be supported by a detailed user documentation e.g., manual or instructions, which describes how the method should be applied.		NARA
		The method contains user documentation that provides a detailed step-by-step procedure for all steps in the derivation of an HEP.	X	Justification NARA provides a User Manual which gives a detailed step-by-step procedure for all aspects of NARA application (GTT selection, EPC selection and anchoring, quantification, extended time factor consideration, and documenting the analysis.
		The method contains user documentation that provides a high level description of how it is applied to derive HEPs, but not all elements of the method are detailed as step-by-step procedures.		
		The method provides only a high level description of its method of application and or data tables for the derivation of HEPs.		

Usability	Desirable	**Attribute 19** **Use of limiting values** The method should provide limiting values. (Relevant Good Practice documents discuss limiting values that are used in member countries).		NARA
		The method provides limiting values and advice on their application.	X	Justification Human Performance Limiting Values are prescribed in the User Manual and their detailed consideration and application further explained in the Technical Basis document (Appendix J).
		The method provides advice on the need to limit claims on human performance but does not provide specific limiting values.		
		The method does not consider the use of limiting values.		

Usability	Indifferent/Essential	**Attribute 20** **Resources** A comparative estimate of the resources (time, cost, utility demands, level of specialist training required, level and type of knowledge required) needed to apply the method in comparison with other HRA methods.		NARA
		The estimated cost of and time required for applying the HRA method is less than or comparable with that of other HRA methods**.**	X	Justification The HEP quantification is quick relative to other techniques. As with any HRA technique, the real effort occurs in the qualitative analysis underpinning the HRA and the time required for this should be comparable with that for the application of other techniques. NARA currently has a mandatory 1.5 day training course to be completed by any assessor wishing to use the technique.
		The estimated cost of and time required for applying the HRA method is in excess of that required for application of other HRA methods**.**		

A3.3 Developer's comments on method evaluation – HCR/ORE & CBDT: a summary of the EPRI HRA Methodology

This OECD HRA report ("Establishing Desirable Attributes of Current Human Reliability Assessment (HRA) Techniques in Nuclear Risk Assessment") provides an evaluation of some of the quantification models contained within the EPRI HRA Methodology (specifically, the CBDT and HCR/ORE methods). The EPRI HRA Methodology is not only a quantification approach; it is an overall framework or approach to conducting HRA. The purpose of this appendix is to summarize the EPRI HRA Methodology, and to indicate those portions of the methodology addressed by this OECD HRA report.

The development of EPRI HRA framework originated from the work published in SHARP and SHARP1[1]. These reference documents were used by a joint industry and US NRC team to develop the supporting HRA requirements of the ASME/ANS PRA Standard[2]. In 2001 the EPRI HRA USER Group was formed to establish a formalized, consistent, consensus approach to HRA which meets the supporting HRA requirements of the ASME/ANS PRA Standard[3].

The EPRI HRA Users Group provides technical support, written guidelines, a living knowledgebase, the software tool (the EPRI HRA Calculator®[4]), training and regular user group meetings to share insights on the development of HRA models. Through this process, the EPRI HRA Methodology has evolved and been refined, and gained community acceptance in the US (and has an increasing international user based). As gaps have been identified, EPRI has done focused research projects to extend and augment the methods to applications beyond Level 1, internal events PRA. For example, EPRI participated in a joint research project with NRC to developed NUREG-1921/EPRI 1023001[5] – Fire HRA. This guidance document includes a specific appendix on how to use the EPRI HRA Methodology to evaluate human failure events (HFEs) in a fire context. Additionally, through the PRA Peer Review process in the US, many applications of the EPRI HRA Methodology have been peer reviewed and accepted.

The overall EPRI HRA methodology consists of the following high level process as outlined in the ASME/ANS PRA Standard which apply to pre-initiator and post-initiator operator actions.

1. Identification.
2. Definition/qualitative analysis.
3. Quantification.
4. Dependency analysis.
5. Uncertainty.

---

1. "Systematic Human Action Reliability Procedure" (SHARP), 1984, NP-3583; and "SHARP1 - A Revised Systematic Human Action Reliability Procedure", 1990, NP-7183-SL, Electric Power Research Institute.
2. ASME/ANS RA-Sa-2009, Addenda to ASME/ANS RA-S–2008, *Standard for Level 1/Large Early Release Frequency Probabilistic Risk Assessment for Nuclear Power Plant Applications*, The American Society of Mechanical Engineers, New York, NY, February 2009.
3. ASME/ANS RA-Sa-2009, Addenda to ASME/ANS RA-S–2008, *Standard for Level 1/Large Early Release Frequency Probabilistic Risk Assessment for Nuclear Power Plant Applications*, The American Society of Mechanical Engineers, New York, NY, February 2009.
4. EPRI HRA Calculator® Version 5.1, EPRI Software Product ID 3002004030, by Scientech a Curtiss Wright Flow Control Company, 2014.
5. *EPRI/NRC-RES Fire Human Reliability Analysis Guidelines.* EPRI, Palo Alto, CA, and U.S. Nuclear Regulatory Commission, Washington, D.C.: 2012. 1023001/NUREG-1921.

Identification

Operator actions are identified by a systematic review of the relevant plant-specific procedures. For pre-initiators this consists of surveillance test and periodic maintenance procedures and for post-initiator operator actions this consists of emergency and abnormal operating procedures in conjunction with a review of the event and fault trees. For each initiating event considered in the PRA, the applicable emergency operating procedures (EOPs), abnormal operating procedures (AOPs), annunciator response procedures etc. are reviewed to identify all operator actions necessary for success. The post-initiator actions may be actions required to initiate (for those systems not automatically initiated), operate, control, isolate, or terminate those systems and components used in preventing or mitigating core damage as defined by the success criteria. In addition to the procedure review, a review of the PRA model is preformed to ensure operators actions included in the PRA as recovery actions are also identified.

Definition/Qualitative analysis

For each identified operator action, the definition consists of identifying the tasks needed to accomplish the operator action (or fail to produce a human failure event) and the PRA context in which the tasks are conducted. For post-initiator operator actions this includes synchronizing the operator actions with the unavailability of functions, systems or components at an appropriate level of detail in the accident sequence and system models. Failures to correctly perform several responses may be grouped into one HFE if the impact of the failures is similar or can be conservatively bounded.

For each HFE, the following are qualitatively addressed and these issues along with operator interviews and simulator observations the definition and qualitative analysis is derived:

- The accident sequence-specific timeline (time available, time required, manipulation time).
- The accident sequence-specific procedural guidance (e.g., AOPs and EOPs).
- The availability of cues and other indications for detection and evaluation of failures and corrective action.
- Degree of clarity of the cues/indications.
- The necessary tasks required for success of the action.
- Quality [type (classroom or simulator) and frequency] of the operator training or experience.
- Quality of the written procedures and administrative controls.
- Human-machine interface.
- Complexity of the required response.
- Environment (e.g., lighting, heat, radiation) under which the operator is working.
- Accessibility of the equipment requiring manipulation.
- Necessity, adequacy, and availability of special tools, parts, clothing, etc.

Quantification

The EPRI HRA Users group has developed the EPRI HRA Calculator® software tool which can be used for documentation of the identification and definition as well as quantifications. The methods included in the software package for post-initiator quantification are:

- HCR/ORE and CBDTM[6].
- THERP[7].

6. *An Approach to the Analysis of Operator Actions in Probabilistic Risk Assessment*, EPRI, Palo Alto, CA: 1992. TR-100259.
7. U.S. Nuclear Regulatory Commission. NUREG/CR-1278, *Handbook of Human Reliability Analysis with Emphasis on Nuclear Power Plant Applications (THERP)*, A. D. Swain and H. E. Guttman, August 1983.

- ASEP[8].
- SPAR-H[9].

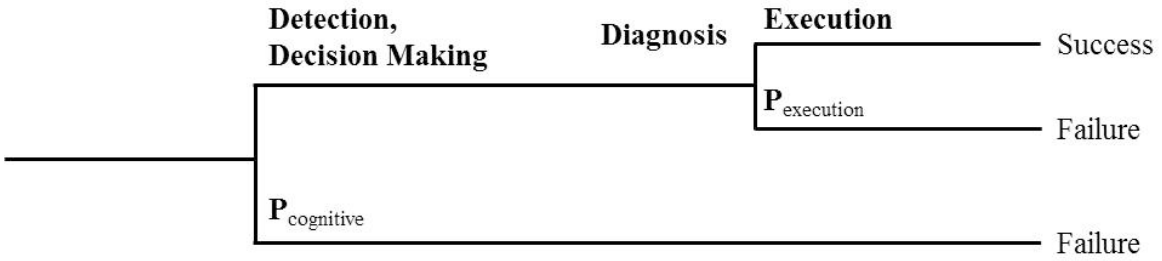
The EPRI recommended methods for quantification of post-initiators are CBDTM, HCR/ORE and THERP. The other methods implemented within the software allow for comparison of the HEPs among different methods.

Following the EPRI approach, post-initiators are evaluated by examining the human failure event as two parts, cognitive errors and execution errors (see figure below). The cognitive contribution is evaluated using the maximum of the CBDTM and HCR/ORE models and the technical basis is described in EPRI TR-100259.

The HCR/ORE model was derived from simulator observations. These simulator observations showed that the HEP can be modeled as function of normalised time and type of crew response structure. Other typical performance shaping factors such as (training, stress, workload, etc.) are amalgamated within the data. The HCR/ORE showed that the HEP is a logarithmic function of time and has more time available increases the HEP rapidly approaches zero. These long-time frame scenarios are not easily observable in the simulator and the failure probability is not dominated by time.

The cause based decision trees (CBDTs) method was developed to provide a floor value to the HCR/ORE which time is not the dominate factor. The CBDTs were also derived from simulator observations into those failure modes and failure mechanisms that challenged operator response. The CBDTM method examines a variety of performance shaping factors relevant to each failure mechanism associated with cognitive performance, and credits immediate recovery opportunities (e.g., self-check, extra crew or STA check) if there is sufficient time available. Both HCR/ORE and CBDTM are geared towards procedure-directed operator actions.

Execution is quantified following THERP. THERP evaluates a variety of execution performance shaping factors and provides the execution failure probability based on the type of execution action. Taken together, the three models provide coverage of a range of cognitive and execution related performance shaping factors, including time.



Dependency Analysis

The EPRI HRA Methodology also includes a systematic approach to performing a dependency analysis. While many dependency issues are typically identified and addressed during the identification and quantification of individual human failure events, the EPRI HRA Methodology also identifies cutsets containing multiple HFEs which appear in the same cutset. The dependency analysis is performed as part

8. U.S. Nuclear Regulatory Commission. NUREG/CR-4772, *Accident Sequence Evaluation Program Human Reliability Analysis Procedure*, February 1987.
9. The SPAR-H Human Reliability Analysis Method (SPAR-H), 2005, Sandia, NUREG/CR-6883.

of the PRA quantification task as required by supporting requirements QU-C1 and QU-C2 of the ASME/ANS PRA Standard. The process includes identification of combinations of HFEs which occur in the cutsets, a systematic evaluation of dependence for the combination, and an approach for implementation of the results into the PRA model. Fundamental to this process is that dependencies between HFE need to be addressed before cutsets are truncated to prevent inappropriate truncation of cutsets containing dependent HFEs. This process is a risk-informed iterative approach and the EPRI HRA Calculator® contains a dependency module which automates much of the process. This method is described in EPRI HRA Calculator® User's Manual[10] as well as several conference papers[11,12,13].

Uncertainty

The EPRI HRA Methodology supports both the evaluation of parametric data uncertainty and modeling uncertainty. For each HEP quantified the EPRI approach adopts the THERP recommendations for the application of error factors based on the overall HEP for the parametric data uncertainty. No mathematical error propagation is recommended. Various, user defined, sensitivity cases can be quantified using the HRA Calculator® in order to evaluate sources of modeling uncertainty.

Conclusion

In conclusion, the EPRI HRA Methodology is not simply the HCR/ORE or CBDTM quantification method. Instead, the EPRI HRA Methodology is an overall process with a software package to assist in HRA development, and when considered holistically the approach collectively addresses all 20 HRA attributes identified in this report. ("Establishing Desirable Attributes of Current Human Reliability Assessment (HRA) Techniques in Nuclear Risk Assessment"). For more information on the EPRI HRA Methodology please contact Mary Presley at mpresley@epri.com or Jeff Julius at jjulius@curtisswright.com.

10. EPRI HRA Calculator® Version 5.1, EPRI Software Product ID 3002004030, by Scientech a Curtiss Wright Flow Control Company, 2014.
11. J. F. Grobbelaar, J. A. Julius, M. Averett, F. Rahn, *Automated Human Reliability Analysis Using the EPRI HRA Calculator®*, ANS PSA 2008 Topical Meeting – Challenges to PSA during the nuclear renaissance, Knoxville, Tennessee, September 7-11, 2008.
12. J. F. Grobbelaar, J. A. Julius, F. Rahn, *Analysis of Dependent Human Failure Events Using the EPRI HRA Calculator®*, International Topical Meeting on Probabilistic Safety Assessment, PSA '05, San Francisco, California, September 11 to 15, 2005.
13. J. F. Grobbelaar, M. Hirt, M, Presley, Human Reliability Dependency Analysis and Model Integration Process, Probabilistic Safety Assessment Meeting PSAM-12, Honolulu, Hawaii, June 23-27, 2014.